

EP1377684A1

Publication Title:

METHODS FOR IDENTIFYING SMALL MOLECULES THAT BIND SPECIFIC
RNA STRUCTURAL MOTIFS

Abstract:

Abstract not available for EP 1377684

(A1) Abstract of corresponding document: WO 02083953

(A1) The present invention relates to a method for screening and identifying test compounds that bind to a preselected target ribonucleic acid ("RNA"). Direct, non-competitive binding assays are advantageously used to screen libraries of compounds for those that selectively bind to a preselected target RNA. Binding of target RNA molecules to a particular test compound is detected using any physical method that measures the altered physical property of the target RNA bound to a test compound. The structure of the test compound attached to the labeled RNA is also determined. The methods used will depend, in part, on the nature of the library screened. The methods of the present invention provide a simple, sensitive assay for high-throughput screening of libraries of compounds to identify pharmaceutical leads.

Courtesy of <http://v3.espacenet.com>

(19) World Intellectual Property Organization
International Bureau(43) International Publication Date
24 October 2002 (24.10.2002)

PCT

(10) International Publication Number
WO 02/083953 A1

- (51) International Patent Classification⁷: C12Q 1/68, C07H 21/02, G01N 27/26
- (21) International Application Number: PCT/US02/11757
- (22) International Filing Date: 11 April 2002 (11.04.2002)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data: 60/282,965 11 April 2001 (11.04.2001) US
- (71) Applicant (for all designated States except US): PTC THERAPEUTICS, INC. [US/US]; 100 Corporate Court, Middlesex Business Center, South Plainfield, NJ 07080 (US).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): RANDO, Robert [US/US]; 3 Brown Court, Annandale, NJ 08801 (US). WELCH, Ellen [US/US]; 33 Hollow Brook Road, Calton, NJ 07830 (US).
- (74) Agents: CORUZZI, Laura, A. et al.; Pennie & Edmonds LLP, 1155 Avenue of the Americas, New York, NY 10036 (US).
- (81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GI, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.
- (84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).
- Published:**
— with international search report
— with amended claims
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*



WO 02/083953 A1

(54) Title: METHODS FOR IDENTIFYING SMALL MOLECULES THAT BIND SPECIFIC RNA STRUCTURAL MOTIFS

(57) **Abstract:** The present invention relates to a method for screening and identifying test compounds that bind to a preselected target ribonucleic acid ("RNA"). Direct, non-competitive binding assays are advantageously used to screen libraries of compounds for those that selectively bind to a preselected target RNA. Binding of target RNA molecules to a particular test compound is detected using any physical method that measures the altered physical property of the target RNA bound to a test compound. The structure of the test compound attached to the labeled RNA is also determined. The methods used will depend, in part, on the nature of the library screened. The methods of the present invention provide a simple, sensitive assay for high-throughput screening of libraries of compounds to identify pharmaceutical leads.

METHODS FOR IDENTIFYING SMALL MOLECULES THAT BIND SPECIFIC RNA STRUCTURAL MOTIFS

5 This application claims the benefit of U.S. Provisional Application No.
60/282,965, filed April 11, 2001, which is incorporated herein by reference in its entirety.

1. INTRODUCTION

10 The present invention relates to a method for screening and identifying test
compounds that bind to a preselected target ribonucleic acid ("RNA"). Direct, non-
competitive binding assays are advantageously used to screen libraries of compounds for
those that selectively bind to a preselected target RNA. Binding of target RNA molecules to
a particular test compound is detected using any physical method that measures the altered
physical property of the target RNA bound to a test compound. The methods of the present
15 invention provide a simple, sensitive assay for high-throughput screening of libraries of
compounds to identify pharmaceutical leads.

2. BACKGROUND OF THE INVENTION

Protein-nucleic acid interactions are involved in many cellular functions,
20 including transcription, RNA splicing, mRNA decay, and mRNA translation. Readily
accessible synthetic molecules that can bind with high affinity to specific sequences of
single- or double-stranded nucleic acids have the potential to interfere with these
interactions in a controllable way, making them attractive tools for molecular biology and
medicine. Successful approaches for blocking function of target nucleic acids include using
25 duplex-forming antisense oligonucleotides (Miller, 1996, *Progress in Nucl. Acid Res. &*
Mol. Biol. 52:261-291; Ojwang & Rando, 1999, Achieving antisense inhibition by
oligodeoxynucleotides containing N₇ modified 2'-deoxyguanosine using tumor necrosis
factor receptor type 1, *METHODS: A Companion to Methods in Enzymology* 18:244-251)
and peptide nucleic acids ("PNA") (Nielsen, 1999, *Current Opinion in Biotechnology*
30 10:71-75), which bind to nucleic acids via Watson-Crick base-pairing. Triplex-forming
anti-gene oligonucleotides can also be designed (Ping *et al.*, 1997, *RNA* 3:850-860;
Aggarwal *et al.*, 1996, *Cancer Res.* 56:5156-5164; U.S. Patent No. 5,650,316), as well as
pyrrole-imidazole polyamide oligomers (Gottesfeld *et al.*, 1997, *Nature* 387:202-205; White
et al., 1998, *Nature* 391:468-471), which are specific for the major and minor grooves of a
35 double helix, respectively.

In addition to synthetic nucleic acids (*i.e.*, antisense, ribozymes, and triplex-forming molecules), there are examples of natural products that interfere with deoxyribonucleic acid ("DNA") or RNA processes such as transcription or translation. For example, certain carbohydrate-based host cell factors, calicheamicin oligosaccharides, interfere with the sequence-specific binding of transcription factors to DNA and inhibit transcription *in vivo* (Ho *et al.*, 1994, Proc. Natl. Acad. Sci. USA 91:9203-9207; Liu *et al.*, 1996, Proc. Natl. Acad. Sci. USA 93:940-944). Certain classes of known antibiotics have been characterized and were found to interact with RNA. For example, the antibiotic thiostreptone binds tightly to a 60-mer from ribosomal RNA (Cundliffe *et al.*, 1990, in The Ribosome: Structure, Function & Evolution (Schlessinger *et al.*, eds.) American Society for Microbiology, Washington, D.C. pp. 479-490). Bacterial resistance to various antibiotics often involves methylation at specific rRNA sites (Cundliffe, 1989, Ann. Rev. Microbiol. 43:207-233). Aminoglycosidic aminocyclitol (aminoglycoside) antibiotics and peptide antibiotics are known to inhibit group I intron splicing by binding to specific regions of the RNA (von Ahsen *et al.*, 1991, Nature (London) 353:368-370). Some of these same aminoglycosides have also been found to inhibit hammerhead ribozyme function (Stage *et al.*, 1995, RNA 1:95-101). In addition, certain aminoglycosides and other protein synthesis inhibitors have been found to interact with specific bases in 16S rRNA (Woodcock *et al.*, 1991, EMBO J. 10:3099-3103). An oligonucleotide analog of the 16S rRNA has also been shown to interact with certain aminoglycosides (Purohit *et al.*, 1994, Nature 370:659-662). A molecular basis for hypersensitivity to aminoglycosides has been found to be located in a single base change in mitochondrial rRNA (Hutchin *et al.*, 1993, Nucleic Acids Res. 21:4174-4179). Aminoglycosides have also been shown to inhibit the interaction between specific structural RNA motifs and the corresponding RNA binding protein. Zapp *et al.* (Cell, 1993, 74:969-978) has demonstrated that the aminoglycosides neomycin B, lividomycin A, and tobramycin can block the binding of Rev, a viral regulatory protein required for viral gene expression, to its viral recognition element in the IIB (or RRE) region of HIV RNA. This blockage appears to be the result of competitive binding of the antibiotics directly to the RRE RNA structural motif.

Single stranded sections of RNA can fold into complex tertiary structures consisting of local motifs such as loops, bulges, pseudoknots, guanosine quartets and turns (Chastain & Tinoco, 1991, Progress in Nucleic Acid Res. & Mol. Biol. 41:131-177; Chow & Bogdan, 1997, Chemical Reviews 97:1489-1514; Rando & Hogan, 1998, Biologic activity of guanosine quartet forming oligonucleotides in "Applied Antisense Oligonucleotide Technology" Stein. & Krieg (eds) John Wiley and Sons, New York, pages

335-352). Such structures can be critical to the activity of the nucleic acid and affect functions such as regulation of mRNA transcription, stability, or translation (Weeks & Crothers, 1993, Science 261:1574-1577). The dependence of these functions on the native three-dimensional structural motifs of single-stranded stretches of nucleic acids makes it difficult to identify or design synthetic agents that bind to these motifs using general, simple-to-use sequence-specific recognition rules for the formation of double- and triple-helical nucleic acids used in the design of antisense and ribozyme type molecules. Approaches to screening generally involve competitive assays designed to identify compounds that disrupt the interaction between a target RNA and a physiological, host cell factor(s) that had been previously identified to specifically interact with that particular target RNA. In general, such assays require the identification and characterization of the host cell factor(s) deemed to be required for the function of the target RNA. Both the target RNA and its preselected host cell binding partner are used in a competitive format to identify compounds that disrupt or interfere with the two components in the assay.

Citation or identification of any reference in Section 2 of this application is not an admission that such reference is available as prior art to the present invention.

3. SUMMARY OF THE INVENTION

The present invention relates to methods for identifying compounds that bind to preselected target elements of nucleic acids including, but not limited to, specific RNA sequences, RNA structural motifs, and/or RNA structural elements. The specific target RNA sequences, RNA structural motifs, and/or RNA structural elements are used as targets for screening small molecules and identifying those that directly bind these specific sequences, motifs, and/or structural elements. For example, methods are described in which a preselected target RNA having a detectable label is used to screen a library of test compounds, preferably under physiologic conditions. Any complexes formed between the target RNA and a member of the library are identified using physical methods that detect the altered physical property of the target RNA bound to a test compound. In particular, the present invention relates to methods for using a target RNA having a detectable label to screen a library of test compounds free in solution, in labeled tubes or microtiter plate, or in a microarray. Compounds in the library that bind to the labeled target RNA will form a detectably labeled complex. The detectably labeled complex can then be identified and removed from the uncomplexed, unlabeled test compounds in the library, and from uncomplexed, labeled target RNA, by a variety of methods, including but not limited to, methods that differentiate changes in the electrophoretic, chromatographic, or thermostable

properties of the complexed target RNA. Such methods include, but are not limited to, electrophoresis, fluorescence spectroscopy, surface plasmon resonance, mass spectrometry, scintillation, proximity assay, structure-activity relationships ("SAR") by NMR spectroscopy, size exclusion chromatography, affinity chromatography, and nanoparticle aggregation. The structure of the test compound attached to the labeled RNA is then determined. The methods used will depend, in part, on the nature of the library screened. For example, assays or microarrays of test compounds, each having an address or identifier, may be deconvoluted, *e.g.*, by cross-referencing the positive sample to original compound list that was applied to the individual test assays. Another method for identifying test compounds includes *de novo* structure determination of the test compounds using mass spectrometry or nuclear magnetic resonance ("NMR"). The test compounds identified are useful for any purpose to which a binding reaction may be put, for example in assay methods, diagnostic procedures, cell sorting, as inhibitors of target molecule function, as probes, as sequestering agents and the like. In addition, small organic molecules which interact specifically with target RNA molecules may be useful as lead compounds for the development of therapeutic agents.

The methods described herein for the identification of compounds that directly bind to a particular preselected target RNA are well suited for high-throughput screening. The direct binding method of the invention offers advantages over drug screening systems for competitors that inhibit the formation of naturally-occurring RNA binding protein:target RNA complexes; *i.e.*, competitive assays. The direct binding method of the invention is rapid and can be set up to be readily performed, *e.g.*, by a technician, making it amenable to high throughput screening. The method of the invention also eliminates the bias inherent in the competitive drug screening systems, which require the use of a preselected host cell factor that may not have physiological relevance to the activity of the target RNA. Instead, the methods of the invention are used to identify any compound that can directly bind to specific target RNA sequences, RNA structural motifs, and/or RNA structural elements, preferably under physiologic conditions. As a result, the compounds so identified can inhibit the interaction of the target RNA with any one or more of the native host cell factors (whether known or unknown) required for activity of the RNA *in vivo*.

The present invention may be understood more fully by reference to the detailed description and examples, which are intended to illustrate non-limiting embodiments of the invention.

3.1. Definitions

As used herein, a "target nucleic acid" refers to RNA, DNA, or a chemically modified variant thereof. In a preferred embodiment, the target nucleic acid is RNA. A target nucleic acid also refers to tertiary structures of the nucleic acids, such as, but not limited to loops, bulges, pseudoknots, guanosine quartets and turns. A target nucleic acid also refers to RNA elements such as, but not limited to, the HIV TAR element, internal ribosome entry site, "slippery site", instability elements, and adenylate uridyate-rich elements, which are described in Section 5.1. Non-limiting examples of target nucleic acids are presented in Section 5.1 and Section 6.

As used herein, a "library" refers to a plurality of test compounds with which a target nucleic acid molecule is contacted. A library can be a combinatorial library, *e.g.*, a collection of test compounds synthesized using combinatorial chemistry techniques, or a collection of unique chemicals of low molecular weight (less than 1000 daltons) that each occupy a unique three-dimensional space.

As used herein, a "label" or "detectable label" is a composition that is detectable, either directly or indirectly, by spectroscopic, photochemical, biochemical, immunochemical, or chemical means. For example, useful labels include radioactive isotopes (*e.g.*, ^{32}P , ^{35}S , and ^3H), dyes, fluorescent dyes, electron-dense reagents, enzymes and their substrates (*e.g.*, as commonly used in enzyme-linked immunoassays, *e.g.*, alkaline phosphatase and horse radish peroxidase), biotin-streptavidin, digoxigenin, or haptens and proteins for which antisera or monoclonal antibodies are available. Moreover, a label or detectable moiety can include a "affinity tag" that, when coupled with the target nucleic acid and incubated with a test compound or compound library, allows for the affinity capture of the target nucleic acid along with molecules bound to the target nucleic acid. One skilled in the art will appreciate that a affinity tag bound to the target nucleic acids has, by definition, a complimentary ligand coupled to a solid support that allows for its capture. For example, useful affinity tags and complimentary partners include, but are not limited to, biotin-streptavidin, complimentary nucleic acid fragments (*e.g.*, oligo dT-oligo dA, oligo T-oligo A, oligo dG-oligo dC, oligo G-oligo C), aptamers, or haptens and proteins for which antisera or monoclonal antibodies are available. The label or detectable moiety is typically bound, either covalently, through a linker or chemical bound, or through ionic, van der Waals or hydrogen bonds to the molecule to be detected.

As used herein, a "dye" refers to a molecule that, when exposed to radiation, emits radiation at a level that is detectable visually or via conventional spectroscopic means.

As used herein, a "visible dye" refers to a molecule having a chromophore that absorbs radiation in the visible region of the spectrum (*i.e.*, having a wavelength of between about 400 nm and about 700 nm) such that the transmitted radiation is in the visible region and can be detected either visually or by conventional spectroscopic means. As used herein, an "ultraviolet dye" refers to a molecule having a chromophore that absorbs radiation in the ultraviolet region of the spectrum (*i.e.*, having a wavelength of between about 30 nm and about 400 nm). As used herein, an "infrared dye" refers to a molecule having a chromophore that absorbs radiation in the infrared region of the spectrum (*i.e.*, having a wavelength between about 700 nm and about 3,000 nm). A "chromophore" is the network of atoms of the dye that, when exposed to radiation, emits radiation at a level that is detectable visually or via conventional spectroscopic means. One of skill in the art will readily appreciate that although a dye absorbs radiation in one region of the spectrum, it may emit radiation in another region of the spectrum. For example, an ultraviolet dye may emit radiation in the visible region of the spectrum. One of skill in the art will also readily appreciate that a dye can transmit radiation or can emit radiation via fluorescence or phosphorescence.

The phrase "pharmaceutically acceptable salt(s)," as used herein includes but is not limited to salts of acidic or basic groups that may be present in test compounds identified using the methods of the present invention. Test compounds that are basic in nature are capable of forming a wide variety of salts with various inorganic and organic acids. The acids that can be used to prepare pharmaceutically acceptable acid addition salts of such basic compounds are those that form non-toxic acid addition salts, *i.e.*, salts containing pharmacologically acceptable anions, including but not limited to sulfuric, citric, maleic, acetic, oxalic, hydrochloride, hydrobromide, hydroiodide, nitrate, sulfate, bisulfate, phosphate, acid phosphate, isonicotinate, acetate, lactate, salicylate, citrate, acid citrate, tartrate, oleate, tannate, pantothenate, bitartrate, ascorbate, succinate, maleate, gentisinate, fumarate, gluconate, glucuronate, saccharate, formate, benzoate, glutamate, methanesulfonate, ethanesulfonate, benzenesulfonate, p-toluenesulfonate and pamoate (*i.e.*, 1,1'-methylene-bis-(2-hydroxy-3-naphthoate)) salts. Test compounds that include an amino moiety may form pharmaceutically or cosmetically acceptable salts with various amino acids, in addition to the acids mentioned above. Test compounds that are acidic in nature are capable of forming base salts with various pharmacologically or cosmetically acceptable cations. Examples of such salts include alkali metal or alkaline earth metal salts and, particularly, calcium, magnesium, sodium lithium, zinc, potassium, and iron salts.

By "substantially one type of test compound," as used herein, is meant that the assay can be performed in such a fashion that at some point, only one compound need be used in each reaction so that, if the result is indicative of a binding event occurring between the target RNA molecule and the test compound, the test compound can be easily identified.

4. DESCRIPTION OF DRAWINGS

FIG. 1. Gel retardation analysis to detect peptide-RNA interactions. In 20 μ l reactions containing increasing concentrations of Tat₄₇₋₅₅ peptide (0.1 μ M, 0.2 μ M, 0.4 μ M, 0.8 μ M, 1.6 μ M) 50 pmole TAR RNA oligonucleotide was added in TK buffer. The reaction mixture was then heated at 90°C for 2 min and allowed to cool slowly to 24°C. 10 ml of 30% glycerol was added to each sample and applied to a 12% non-denaturing polyacrylamide gel. The gel was electrophoresed using 1200 volt-hours at 4°C in TBE Buffer. Following electrophoresis, the gel was dried and the radioactivity was quantitated with a phosphorimager. The concentration of peptide added is indicated above each lane.

FIG. 2. Gentamicin interacts with an oligonucleotide corresponding to the 16S rRNA. 20 μ l reactions containing increasing concentrations of gentamicin (1 ng/ml, 10 ng/ml, 100 ng/ml, 1 μ g/ml, 10 μ g/ml, 50 μ g/ml, 500 μ g/ml) were added to 50 pmole RNA oligonucleotide in TKM buffer, heated at 90°C for 2 min and allowed to cool slowly to 24°C. Then 10 μ l of 30% glycerol was added to each sample and the samples were applied to a 13.5% non-denaturing polyacrylamide gel. The gel was electrophoresed using 1200 volt-hours at 4°C in TBE Buffer. Following electrophoresis, the gel was dried and the radioactivity was quantitated using a phosphorimager. The concentration of gentamicin added is indicated above each lane.

FIG. 3. The presence of 10 pg/ml gentamicin produces a gel mobility shift in the presence of the 16S rRNA oligonucleotide. 20 μ l reactions containing increasing concentrations of gentamicin (100 ng/ml, 10 ng/ml, 1 ng/ml, 100 pg/ml, and 10 pg/ml) were added to 50 pmole RNA oligonucleotide in TKM buffer were treated as described for Figure 2.

FIG. 4. Gentamicin binding to the 16S rRNA oligonucleotide is weak in the absence of MgCl₂. Reaction mixtures containing gentamicin (1 mg/ml, 100 μ g/ml,

10 µg/ml, 1 µg/ml, 0.1 µg/ml, and 10 ng/ml) were treated as described in Figure 2 except that the TKM buffer does not contain MgCl₂.

FIG. 5. Gel retardation analysis to detect peptide-RNA interactions. In reactions containing increasing concentrations of Tat₄₇₋₅₈ peptide (0.1 µM, 0.2 µM, 0.4 µM, 0.8 µM, 1.6 µM) 50 pmole TAR RNA oligonucleotide was added in TK buffer. The reaction mixture was then heated at 90°C for 2 min and allowed to cool slowly to 24°C. The reactions were loaded onto a SCE9610 automated capillary electrophoresis apparatus (SpectruMedix; State College, Pennsylvania). The peaks correspond to the amount of free TAR RNA ("TAR") or the Tat-TAR complex ("Tat-TAR"). The concentration of peptide added is indicated below each lane.

5. DETAILED DESCRIPTION OF THE INVENTION

The present invention relates to methods for identifying compounds that bind to preselected target elements of nucleic acids, in particular, RNAs, including but not limited to preselected target RNA sequencing structural motifs, or structural elements. Methods are described in which a preselected target RNA having a detectable label is used to screen a library of test compounds. Any complexes formed between the target RNA and a member of the library are identified using physical methods that detect the altered physical property of the target RNA bound to a test compound. Changes in the physical property of the RNA-test compound complex relative to the target RNA or test compound can be measured by methods such as, but not limited to, methods that detect a change in mobility due to a change in mass, change in charge, or a change in thermostability. Such methods include, but are not limited to, electrophoresis, fluorescence spectroscopy, surface plasmon resonance, mass spectrometry, scintillation, proximity assay, structure-activity relationships ("SAR") by NMR spectroscopy, size exclusion chromatography, affinity chromatography, and nanoparticle aggregation. In particular, the present invention relates to methods for using a target RNA having a detectable label to screen a library of test compounds free in solution, in labeled tubes or microtiter plate, or in a microarray. Compounds in the library that bind to the labeled target RNA will form a detectably labeled complex. The detectably labeled complex can then be identified and removed from the unlabeled, uncomplexed test compounds in the library by a variety of methods capable of differentiating changes in the physical properties of the complexed target RNA. The structure of the test compound attached to the labeled RNA is also determined. The methods used will depend, in part, on the nature of the library screened. For example, assays or microarrays of test compounds,

each having an address or identifier, may be deconvoluted, e.g., by cross-referencing the positive sample to an original compound list that was applied to the individual test assays. Another method for identifying test compounds includes *de novo* structure determination of the test compounds using mass spectrometry or nuclear magnetic resonance ("NMR").

Thus, the methods of the present invention provide a simple, sensitive assay for high-throughput screening of libraries of test compounds, in which the test compounds of the library that specifically bind a preselected target nucleic acid are easily distinguished from non-binding members of the library. The structures of the binding molecules are deciphered from the input library by methods depending on the type of library that is used. The test compounds so identified are useful for any purpose to which a binding reaction may be put, for example in assay methods, diagnostic procedures, cell sorting, as inhibitors of target molecule function, as probes, as sequestering agents and lead compounds for development of therapeutics, and the like. Small organic compounds that are identified to interact specifically with the target RNA molecules are particularly attractive candidates as lead compounds for the development of therapeutic agents.

The assay of the invention reduces bias introduced by competitive binding assays which require the identification and use of a host cell factor (presumably essential for modulating RNA function) as a binding partner for the target RNA. The assays of the present invention are designed to detect any compound or agent that binds to the target RNA, preferably under physiologic conditions. Such agents can then be tested for biological activity, without establishing or guessing which host cell factor or factors is required for modulating the function and/or activity of the target RNA.

Section 5.1 describes examples of protein-RNA interactions that are important in a variety of cellular functions and several target RNA elements that can be used to identify test compounds. Compounds that inhibit these interactions by binding to the RNA and successfully competing with the natural protein or host cell factor that endogenously binds to the RNA may be important, e.g., in treating or preventing a disease or abnormal condition, such as an infection or unchecked growth. Section 5.2 describes detectable labels for target nucleic acids that are useful in the methods of the invention. Section 5.3 describes libraries of test compounds. Section 5.4 provides conditions for binding a labeled target RNA to a test compound of a library and detecting RNA binding to a test compound using the methods of the invention. Section 5.5 provides methods for separating complexes of target RNAs bound to a test compound from an unbound RNA. Section 5.6 describes methods for identifying test compounds that are bound to the target RNA. Section 5.7 describes a secondary, biological screen of test compounds identified by

the methods of the invention to test the effect of the test compounds *in vivo*. Section 5.8 describes the use of test compounds identified by the methods of the invention for treating or preventing a disease or abnormal condition in mammals.

5

5.1. Biologically Important RNA-Host Cell Factor Interactions

Nucleic acids, and in particular RNAs, are capable of folding into complex tertiary structures that include bulges, loops, triple helices and pseudoknots, which can provide binding sites for host cell factors, such as proteins and other RNAs. RNA-protein and RNA-RNA interactions are important in a variety cellular functions, including transcription, RNA splicing, RNA stability and translation. Furthermore, the binding of such host cell factors to RNAs may alter the stability and translational efficiency of such RNAs, and according affect subsequent translation. For example, some diseases are associated with protein overproduction or decreased protein function. In this case, the identification of compounds to modulate RNA stability and translational efficiency will be useful to treat and prevent such diseases.

The methods of the present invention are useful for identifying test compounds that bind to target RNA elements in a high throughput screening assay of libraries of test compounds in solution. In particular, the methods of the present invention are useful for identifying a test compound that binds to a target RNA elements and inhibits the interaction of that RNA with one or more host cell factors *in vivo*. The molecules identified using the methods of the invention are useful for inhibiting the formation of a specific bound RNA:host cell factor complexes *in vivo*.

In some embodiments, test compounds identified by the methods of the invention are useful for increasing or decreasing the translation of messenger RNAs ("mRNAs"), e.g., protein production, by binding to one or more regulatory elements in the 5' untranslated region, the 3' untranslated region, or the coding region of the mRNA. Compounds that bind to mRNA can, *inter alia*, increase or decrease the rate of mRNA processing, alter its transport through the cell, prevent or enhance binding of the mRNA to ribosomes, suppressor proteins or enhancer proteins, or alter mRNA stability. Accordingly, compounds that increase or decrease mRNA translation can be used to treat or prevent disease. For example, diseases associated with protein overproduction, such as amyloidosis, or with the production of mutant proteins, such as *Ras*, can be treated or prevented by decreasing translation of the mRNA that codes for the overproduced protein, thus inhibiting production of the protein. Conversely, the symptoms of diseases associated with decreased protein function, such as hemophilia, may be treated by increasing

translation of mRNA coding for the protein whose function is decreased, *e.g.*, factor IX in some forms of hemophilia.

The methods of the invention can be used to identify compounds that bind to mRNAs coding for a variety of proteins with which the progression of diseases in mammals is associated. These mRNAs include, but are not limited to, those coding for amyloid protein and amyloid precursor protein; anti-angiogenic proteins such as angiostatin, endostatin, METH-1 and METH-2; apoptosis inhibitor proteins such as survivin, clotting factors such as Factor IX, Factor VIII, and others in the clotting cascade; collagens; cyclins and cyclin inhibitors, such as cyclin dependent kinases, cyclin D1, cyclin E, WAF1, cdk4 inhibitor, and MTS1; cystic fibrosis transmembrane conductance regulator gene (CFTR); cytokines such as IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, IL-13, IL-14, IL-15, IL-16, IL-17 and other interleukins; hematopoietic growth factors such as erythropoietin (Epo); colony stimulating factors such as G-CSF, GM-CSF, M-CSF, SCF and thrombopoietin; growth factors such as BNDF, BMP, GGRP, EGF, FGF, GDNF, GGF, HGF, IGF-1, IGF-2, KGF, myotrophin, NGF, OSM, PDGF, somatotrophin, TGF- β , TGF- α and VEGF; antiviral cytokines such as interferons, antiviral proteins induced by interferons, TNF- α , and TNF- β ; enzymes such as cathepsin K, cytochrome P-450 and other cytochromes, farnesyl transferase, glutathione-S transferases, heparanase, HMG CoA synthetase, N-acetyltransferase, phenylalanine hydroxylase, phosphodiesterase, ras carboxyl-terminal protease, telomerase and TNF converting enzyme; glycoproteins such as cadherins, *e.g.*, N-cadherin and E-cadherin; cell adhesion molecules; selectins; transmembrane glycoproteins such as CD40; heat shock proteins; hormones such as 5- α reductase, atrial natriuretic factor, calcitonin, corticotrophin releasing factor, diuretic hormones, glucagon, gonadotropin, gonadotropin releasing hormone, growth hormone, growth hormone releasing factor, somatotrophin, insulin, leptin, luteinizing hormone, luteinizing hormone releasing hormone, parathyroid hormone, thyroid hormone, and thyroid stimulating hormone; proteins involved in immune responses, including antibodies, CTLA4, hemagglutinin, MHC proteins, VLA-4, and kallikrein-kininogen-kinin system; ligands such as CD4; oncogene products such as *sis*, *hst*, protein tyrosine kinase receptors, *ras*, *abl*, *mos*, *myc*, *fos*, *jun*, *H-ras*, *ki-ras*, *c-fms*, *bcl-2*, *L-myc*, *c-myc*, *gip*, *gsp*, and *HER-2*; receptors such as bombesin receptor, estrogen receptor, GABA receptors, growth factor receptors including EGFR, PDGFR, FGFR, and NGFR, GTP-binding regulatory proteins, interleukin receptors, ion channel receptors, leukotriene receptor antagonists, lipoprotein receptors, opioid pain receptors, substance P receptors, retinoic acid and retinoid receptors, steroid receptors, T-cell receptors, thyroid hormone receptors, TNF receptors; tissue

plasminogen activator; transmembrane receptors; transmembrane transporting systems, such as calcium pump, proton pump, Na/Ca exchanger, MRP1, MRP2, P170, LRP, and cMOAT; transferrin; and tumor suppressor gene products such as *APC*, *brca1*, *brca2*, *DCC*, *MCC*, *MTS1*, *NF1*, *NF2*, *nm23*, *p53* and *Rb*. In addition to the eukaryotic genes listed above, the invention, as described, can be used to define molecules that interrupt viral, bacterial or fungal transcription or translation efficiencies and therefore form the basis for a novel anti-infectious disease therapeutic. Other target genes include, but are not limited to, those disclosed in Section 5.1 and Section 6.

The methods of the invention can be used to identify mRNA-binding test compounds for increasing or decreasing the production of a protein, thus treating or preventing a disease associated with decreasing or increasing the production of said protein, respectively. The methods of the invention may be useful for identifying test compounds for treating or preventing a disease in mammals, including cats, dogs, swine, horses, goats, sheep, cattle, primates and humans. Such diseases include, but are not limited to, amyloidosis, hemophilia, Alzheimer's disease, atherosclerosis, cancer, gigantism, dwarfism, hypothyroidism, hyperthyroidism, inflammation, cystic fibrosis, autoimmune disorders, diabetes, aging, obesity, neurodegenerative disorders, and Parkinson's disease. Other diseases include, but are not limited to, those described in Section 5.1 and diseases caused by aberrant expression of the genes disclosed in Example 6. In addition to the eukaryotic genes listed above, the invention, as described, can be used to define molecules that interrupt viral, bacterial or fungal transcription or translation efficiencies and therefore form the basis for a novel anti-infectious disease therapeutic.

In other embodiments, test compounds identified by the methods of the invention are useful for preventing the interaction of an RNA, such as a transfer RNA ("tRNA"), an enzymatic RNA or a ribosomal RNA ("rRNA"), with a protein or with another RNA, thus preventing, *e.g.*, assembly of an *in vivo* protein-RNA or RNA-RNA complex that is essential for the viability of a cell. The term "enzymatic RNA," as used herein, refers to RNA molecules that are either self-splicing, or that form an enzyme by virtue of their association with one or more proteins, *e.g.*, as in RNase P, telomerase or small nuclear ribonuclear protein particles. For example, inhibition of an interaction between rRNA and one or more ribosomal proteins may inhibit the assembly of ribosomes, rendering a cell incapable of synthesizing proteins. In addition, inhibition of the interaction of precursor rRNA with ribonucleases or ribonucleoprotein complexes (such as RNase P) that process the precursor rRNA prevent maturation of the rRNA and its assembly into ribosomes. Similarly, a tRNA:tRNA synthetase complex may be inhibited by test

compounds identified by the methods of the invention such that tRNA molecules do not become charged with amino acids. Such interactions include, but are not limited to, rRNA interactions with ribosomal proteins, tRNA interactions with tRNA synthetase, RNase P protein interactions with RNase P RNA, and telomerase protein interactions with telomerase RNA.

In other embodiments, test compounds identified by the methods of the invention are useful for treating or preventing a viral, bacterial, protozoan or fungal infection. For example, transcriptional up-regulation of the genes of human immunodeficiency virus type 1 ("HIV-1") requires binding of the HIV Tat protein to the HIV trans-activation response region RNA ("TAR RNA"). HIV TAR RNA is a 59-base stem-loop structure located at the 5'-end of all nascent HIV-1 transcripts (Jones & Peterlin, 1994, *Annu. Rev. Biochem.* 63:717-43). Tat protein is known to interact with uracil 23 in the bulge region of the stem of TAR RNA. Thus, TAR RNA is a potential binding target for test compounds, such as small peptides and peptide analogs that bind to the bulge region of TAR RNA and inhibit formation of a Tat-TAR RNA complex involved in HIV-1 upregulation (see Hwang *et al.*, 1999 *Proc. Natl. Acad. Sci. USA* 96:12997-13002). Accordingly, test compounds that bind to TAR RNA are useful as anti-HIV therapeutics (Hamy *et al.*, 1997, *Proc. Natl. Acad. Sci. USA* 94:3548-3553; Hamy *et al.*, 1998, *Biochemistry* 37:5086-5095; Mei *et al.*, 1998, *Biochemistry* 37:14204-14212), and therefore, are useful for treating or preventing AIDS.

The methods of the invention can be used to identify test compounds to treat or prevent viral, bacterial, protozoan or fungal infections in a patient. In some embodiments, the methods of the invention are useful for identifying compounds that decrease translation of microbial genes by interacting with mRNA, as described above, or for identifying compounds that inhibit the interactions of microbial RNAs with proteins or other ligands that are essential for viability of the virus or microbe. Examples of microbial target RNAs useful in the present invention for identifying antiviral, antibacterial, anti-protozoan and anti-fungal compounds include, but are not limited to, general antiviral and anti-inflammatory targets such as mRNAs of INF α , INF γ , RNase L, RNase L inhibitor protein, PKR, tumor necrosis factor, interleukins 1-15, and IMP dehydrogenase; internal ribosome entry sites; HIV-1 CT rich domain and RNase H mRNA; HCV internal ribosome entry site (required to direct translation of HCV mRNA), and the 3'-untranslated tail of HCV genomes; rotavirus NSP3 binding site, which binds the protein NSP3 that is required for rotavirus mRNA translation; HBV epsilon domain; Dengue virus 5' and 3' untranslated regions, including IRES; INF α , INF β and INF γ ; plasmodium falciparum mRNAs; the 16S

ribosomal subunit ribosomal RNA and the RNA component of RNase P of bacteria; and the RNA component of telomerase in fungi and cancer cells. Other target viral and bacterial mRNAs include, but are not limited to, those disclosed in Section 6.

5 One of skill in the art will appreciate that, although such target RNAs are functionally conserved in various species (e.g., from yeast to humans), they exhibit nucleotide sequence and structural diversity. Therefore, inhibition of, for example, yeast telomerase by an anti-fungal compound identified by the methods of the invention might not interfere with human telomerase and normal human cell proliferation.

10 Thus, the methods of the invention can be used to identify test compounds that interfere with one or more target RNA interactions with host cell factors that are important for cell growth or viability, or essential in the life cycle of a virus, a bacterium, a protozoa or a fungus. Such test compounds and/or congeners that demonstrate desirable biologic and pharmacologic activity can be administered to a patient in need thereof in order to treat or prevent a disease caused by viral, bacterial, protozoan, or fungal infections. Such
15 diseases include, but are not limited to, HIV infection, AIDS, human T-cell leukemia, SIV infection, FIV infection, feline leukemia, hepatitis A, hepatitis B, hepatitis C, Dengue fever, malaria, rotavirus infection, severe acute gastroenteritis, diarrhea, encephalitis, hemorrhagic fever, syphilis, legionella, whooping cough, gonorrhea, sepsis, influenza, pneumonia, tinea infection, candida infection, and meningitis.

20 Non-limiting examples of RNA elements involved in the regulation of gene expression, *i.e.*, mRNA stability, translational efficiency via translational initiation and ribosome assembly, *etc.*, include the HIV TAR element, internal ribosome entry site, "slippery site", instability elements, and adenylate uridylate-rich elements, as discussed below.

5.1.1. HIV TAR Element

Transcriptional up-regulation of the genes of human immunodeficiency virus type 1 ("HIV-1") requires binding of the HIV Tat protein to the HIV trans-activation
30 response region RNA ("TAR RNA"), a 59-base stem-loop structure located at the 5' end of all nascent HIV-1 transcripts (Jones & Peterlin, 1994, Annu. Rev. Biochem. 63:717-43). Tat protein is known to interact with uracil 23 in the bulge region of the stem of TAR RNA. Thus, TAR RNA is a useful binding target for test compounds, such as small peptides and peptide analogs that bind to the bulge region of TAR RNA and inhibit formation of a Tat-TAR RNA complex involved in HIV-1 up-regulation (see Hwang *et al.*, 1999 Proc. Natl.
35 Acad. Sci. USA 96:12997-13002). Accordingly, test compounds that bind to TAR RNA

can be useful as anti-HIV therapeutics (Hamy *et al.*, 1997, Proc. Natl. Acad. Sci. USA 94:3548-3553; Hamy *et al.*, 1998, Biochemistry 37:5086-5095; Mei *et al.*, 1998, Biochemistry 37:14204-14212), and therefore, are useful for treating or preventing AIDS.

5

5.1.2. Internal Ribosome Entry Site ("IRES")

Internal ribosome entry sites ("IRES") are found in the 5' untranslated regions ("5' UTR") of several mRNAs, and are thought to be involved in the regulation of translational efficiency. When the IRES element is present on an mRNA downstream of a translational stop codon, it directs ribosomal re-entry (Ghatts *et al.*, 1991, Mol. Cell. Biol. 11:5848-5959), which permits initiation of translation at the start of a second open reading frame.

As reviewed by Jang *et al.*, a large segment of the 5' nontranslated region, approximately 400 nucleotides in length, promotes internal entry of ribosomes independent of the non-capped 5' end of picornavirus mRNAs (mammalian plus-strand RNA viruses whose genomes serve as mRNA). This 400 nucleotide segment (IRES), maps approximately 200 nt down-stream from the 5' end and is highly structured. IRES elements of different picornaviruses, although functionally similar *in vitro* and *in vivo*, are not identical in sequence or structure. However, IRES elements of the genera entero- and rhinoviruses, on the one hand, and cardio- and aphthoviruses, on the other hand, reveal similarities corresponding to phylogenetic kinship. All IRES elements contain a conserved Yn-Xm-AUG unit (Y, pyrimidine; X, nucleotide) which appears essential for IRES function. The IRES elements of cardio-, entero- and aphthoviruses bind a cellular protein, p57. In the case of cardiopviruses, the interaction between a specific stem-loop of the IRES is essential for translation *in vitro*. The IRES elements of entero- and cardiopviruses also bind the cellular protein, p52, but the significance of this interaction remains to be shown. The function of p57 or p52 in cellular metabolism is unknown. Since picornaviral IRES elements function *in vivo* in the absence of any viral gene products, is speculated that IRES-like elements may also occur in specific cellular mRNAs releasing them from cap-dependent translation (Jang *et al.*, 1990, Enzyme 44(1-4):292-309).

30

5.1.3. "Slippery Site"

Programmed, or directed, ribosomal frameshifting, when ribosomes shift from one translation reading frame to another and synthesize two viral proteins from a single viral mRNA, is directed by a unique site in viral mRNAs called the "slippery site." The slippery site directs ribosomal frameshifting in the -1 or +1 direction that causes the

35

ribosome to slip by one base in the 5' direction thereby placing the ribosome in the new reading frame to produce a new protein.

5 Programmed, or directed, ribosomal frameshifting is of particular value to viruses that package their plus strands, as it eliminates the need to splice their mRNAs and reduces the risk of packaging defective genomes and regulates the ratio of viral proteins synthesized. Examples of programmed translational frameshifting (both +1 and -1 shifts) have been identified in ScV systems (Lopinski *et al.*, 2000, Mol. Cell. Biol. 20(4):1095-103, retroviruses (Falk *et al.*, 1993, J. Virol. 67:273-6277; Jacks & Varmus, 1985, Science 230:1237-1242; Morikawa & Bishop, 1992, Virology 186:389-397; Nam *et al.*, 1993, J. Virol. 67:196-203); coronaviruses (Brierley *et al.*, 1987, EMBO J. 6:3779-3785; Herold & Siddell, 1993, Nucleic Acids Res. 21:5838-5842); giardiaviruses, which are also members of the Totiviridae (Wang *et al.*, 1993, Proc. Natl. Acad. Sci. USA 90:8595-8599); two bacterial genes (Blinkowa & Walker, 1990, Nucleic Acids Res., 18:1725-1729; Craigen & Caskey, 1986, Nature 322:273); bacteriophage genes (Condrón *et al.*, 1991, Nucleic Acids Res. 19:5607-5612); astroviruses (Marczinke *et al.*, 1994, J. Virol. 68:5588-5595); the yeast EST3 gene (Lundblad & Morris, 1997, Curr. Biol. 7:969-976); and the rat, mouse, *Xenopus*, and *Drosophila* ornithine decarboxylase antizymes (Matsufuji *et al.*, 1995, Cell 80:51-60); and a significant number of cellular genes (Herold & Siddell, 1993, Nucleic Acids Res. 21:5838-5842).

20 Drugs targeted to ribosomal frameshifting minimize the problem of virus drug resistance because this strategy targets a host cellular process rather than one introduced into the cell by the virus, which minimizes the ability of viruses to evolve drug-resistant mutants. Compounds that target the RNA elements involved in regulating programmed frameshifting should have several advantages, including (a) any selective pressure on the host cellular translational machinery to adapt to the drugs would have to occur at the host evolutionary time scale, which is on the order of millions of years, (b) ribosomal frameshifting is not used to express any host proteins, and (c) altering viral frameshifting efficiencies by modulating the activity of a host protein minimizing the likelihood that the virus will acquire resistance to such inhibition by mutations in its own genome.

5.1.4. Instability Elements

35 "Instability elements" may be defined as specific sequence elements that promote the recognition of unstable mRNAs by cellular turnover machinery. Instability

elements have been found within mRNA protein coding regions as well as untranslated regions.

Altering the control of stability of normal mRNAs may lead to disease. The alteration of mRNA stability has been implicated in diseases such as, but not limited to, cancer, immune disorders, heart disease, and fibrotic disorders.

There are several examples of mutations that delete instability elements which then result in stabilization of mRNAs that may be involved in the onset of cancer. In Burkitt's lymphoma, a portion of the *c-myc* proto-oncogene is translocated to an Ig locus, producing a form of the *c-myc* mRNA that is five times more stable (*see, e.g.,* Kapstein *et al.*, 1996, J. Biol. Chem. 271(31):18875-84). The highly oncogenic *v-fos* mRNA lacks the 3' UTR adenylylate uridylylate rich element ("ARE") that is found in the more labile and weakly oncogenic *c-fos* mRNA (*see, e.g.,* Schiavi *et al.*, 1992, Biochim Biophys Acta. 1114(2-3):95-106). Differences between the benign cervical lesions brought about by nonintegrated circular human papillomavirus type 16 and its integrated form, that lacks the 3' UTR ARE and correlates with cervical carcinomas, may be a consequence of stabilizing the E6/E7 transcripts encoding oncogenic proteins. Integration of the virus results in deletion of the ARE instability element, resulting in stabilization of the transcripts and over-expression of the proteins (*see, e.g.,* Jeon & Lambert, 1995, Proc. Natl. Acad. Sci. USA 92(5):1654-8). Deletion of AREs from the 3' UTR of the IL-2 and IL-3 genes promotes increased stabilization of these mRNAs, high expression of these proteins, and leads to the formation of cancerous cells (*see, e.g.,* Stoecklin *et al.*, 2000, Mol. Cell. Biol. 20(11):3753-63).

Mutations in trans-acting factors involved in mRNA turnover may also promote cancer. In monocytic tumors, the lymphokine GM-CSF mRNA is specifically stabilized as a consequence of an oncogenic lesion in a trans-acting factor that controls mRNA turnover rates. Furthermore, the normally unstable IL-3 transcript is inappropriately long-lived in most tumor cells. Similarly, the labile GM-CSF mRNA is greatly stabilized in bladder carcinoma cells. *See, e.g.,* Bickel *et al.*, 1990, J. Immunol. 145(3):840-5.

The immune system is regulated by a large number of regulatory molecules that either activate or inhibit the immune response. It has now been clearly demonstrated that stability of the transcripts encoding these proteins are highly regulated. Altered regulation of these molecules leads to mis-regulation of this process and can result in drastic medical consequences. For example, recent results using transgenic mice have shown that mis-regulation of the stability of the important modulator TNF α mRNA leads to diseases

such as, but not limited to, rheumatoid arthritis and a Crohn's-like liver disease. *See, e.g.,* Clark, 2000, *Arthritis Res.* 2(3):172-4.

Smooth muscle in the heart is modulated by the β -adrenergic receptor, which in turn responds to the sympathetic neurotransmitter norepinephrine and the adrenal hormone epinephrine. Chronic heart failure is characterized by impairment of smooth muscle cells, which results, in part, from the more rapid decay of the β -adrenergic receptor mRNA. *See, e.g.,* Ellis & Frielle, 1999, *Biochem. Biophys. Res. Commun.* 258(3):552-8.

A large number of diseases result from over-expression of collagen. For example, cirrhosis results from damage to the liver as a consequence of cancer, viral infection, or alcohol abuse. Such damage causes mis-regulation of collagen expression, leading to the formation of large collagen deposits. Recent results indicate that the sizeable increase in collagen expression is largely attributable to stabilization of its mRNA. *See, e.g.,* Lindquist *et al.*, 2000, *Am. J. Physiol. Gastrointest. Liver Physiol.* 279(3):G471-6.

5.1.5. Adenylate Uridylate-rich Elements ("ARE")

Adenylate uridylate-rich elements ("ARE") are found in the 3' untranslated regions ("3' UTR") of several mRNAs, and involved in the turnover of mRNAs, such as but not limited to transcription factors, cytokines, and lymphokines. AREs may function both as stabilizing and destabilizing elements. ARE mRNAs are classified into five groups, depending on sequence (Bakheet *et al.*, 2001, *Nucl. Acids Res.* 29(1):246-254). An ongoing database at the web site <http://rc.kfshrc.edu.sa/ared> contains ARE-containing mRNAs and their cluster groups, which is incorporated by reference in its entirety. The ARE motifs are classified as follows:

25	Group I Cluster	(AUUUAUUUUUUUUUUUUU)	SEQ ID NO: 1
	Group II Cluster	(AUUUAUUUUUUUUUUU) stretch	SEQ ID NO: 2
	Group III Cluster	(WAUUUAUUUAUUUAUW) stretch	SEQ ID NO: 3
	Group IV Cluster	(WWAUUUUUUUUUWW) stretch	SEQ ID NO: 4
30	Group V Cluster	(WWWWAUUUUUUUWWWW) stretch	SEQ ID NO: 5

The ARE-mRNAs were clustered into five groups containing five, four, three and two pentameric repeats, while the last group contains only one pentamer within the 13-bp ARE pattern. Functional categories were assigned whenever possible according to NCBI-COG functional annotation (Tatusov *et al.*, 2001, *Nucleic Acids Research*, 29(1): 22-28), in addition to the categories: inflammation, immune response, development/differentiation, using an extensive literature search.

Group I contains many secreted proteins including GM-CSF, IL-1, IL-11, IL-12 and Gro- β that affect the growth of hematopoietic and immune cells (Witsell & Schook, 1992, Proc. Natl Acad. Sci. USA, 89:4754-4758). Although TNF α is both a pro-inflammatory and anti-tumor protein, there is experimental evidence that it can act as a growth factor in certain leukemias and lymphomas (Liu *et al.*, 2000, J. Biol. Chem. 275:21086-21093).

Unlike Group I, Groups II-V contain functionally diverse gene families comprising immune response, cell cycle and proliferation, inflammation and coagulation, angiogenesis, metabolism, energy, DNA binding and transcription, nutrient transportation and ionic homeostasis, protein synthesis, cellular biogenesis, signal transduction, and apoptosis (Bakheet *et al.*, 2001, Nucl. Acids Res. 29(1):246-254).

Several groups have described ARE-binding proteins that influence the ARE-mRNA stability. Among the well-characterized proteins are the mammalian homologs of ELAV (embryonic lethal abnormal vision) proteins including AUF1, HuR and He1-N2 (Zhang *et al.*, 1993, Mol. Cell. Biol. 13:7652-7665; Levine *et al.*, 1993, Mol. Cell. Biol. 13:3494-3504; Ma *et al.*, 1996, J. Biol. Chem. 271:8144-8151). The zinc-finger protein tristetraprolin has been identified as another ARE-binding protein with destabilizing activity on TNF α , IL-3 and GM-CSF mRNAs (Stoecklin *et al.*, 2000, Mol. Cell. Biol. 20:3753-3763; Carballo *et al.*, 2000, Blood 95:1891-1899).

Since ARE-containing genes are clearly important in biological systems, including but not limited to a number of the early response genes that regulate cell proliferation and responses to exogenous agents, the identification of compounds that bind to one or more of the ARE clusters and potentially modulate the stability of the target RNA can potentially be of value as a therapeutic.

5.2. Detectably Labeled Target RNAs

Target nucleic acids, including but not limited to RNA and DNA, useful in the methods of the present invention have a label that is detectable via conventional spectroscopic means or radiographic means. Preferably, target nucleic acids are labeled with a covalently attached dye molecule. Useful dye-molecule labels include, but are not limited to, fluorescent dyes, phosphorescent dyes, ultraviolet dyes, infrared dyes, and visible dyes. Preferably, the dye is a visible dye.

Useful labels in the present invention can include, but are not limited to, spectroscopic labels such as fluorescent dyes (*e.g.*, fluorescein and derivatives such as fluorescein isothiocyanate (FITC) and Oregon Green™, rhodamine and derivatives (*e.g.*,

Texas red, tetramethylrhodimine isothiocyanate (TRITC), bora-3a,4a-diaza-s-indacene (BODIPY®) and derivatives, *etc.*), digoxigenin, biotin, phycoerythrin, AMCA, CyDye™, and the like), radiolabels (*e.g.*, ^3H , ^{125}I , ^{35}S , ^{14}C , ^{32}P , ^{33}P , *etc.*), enzymes (*e.g.*, horse radish peroxidase, alkaline phosphatase *etc.*), spectroscopic colorimetric labels such as colloidal gold or colored glass or plastic (*e.g.* polystyrene, polypropylene, latex, *etc.*) beads, or nanoparticles – nanoclusters of inorganic ions with defined dimension from 0.1 to 1000 nm. Useful affinity tags and complimentary partners include, but are not limited to, biotin-streptavidin, complimentary nucleic acid fragments (*e.g.*, oligo dT-oligo dA, oligo T-oligo A, oligo dG-oligo dC, oligo G-oligo C), aptamer-streptavidin, or haptens and proteins for which antisera or monoclonal antibodies are available. The label may be coupled directly or indirectly to a component of the detection assay (*e.g.*, the detection reagent) according to methods well known in the art. A wide variety of labels may be used, with the choice of label depending on sensitivity required, ease of conjugation with the compound, stability requirements, available instrumentation, and disposal provisions.

In one embodiment, nucleic acids that are labeled at one or more specific locations are chemically synthesized using phosphoramidite or other solution or solid-phase methods. Detailed descriptions of the chemistry used to form polynucleotides by the phosphoramidite method are well known (*see, e.g.*, Caruthers *et al.*, U.S. Pat. Nos. 4,458,066 and 4,415,732; Caruthers *et al.*, 1982, Genetic Engineering 4:1-17; *Users Manual Model 392 and 394 Polynucleotide Synthesizers*, 1990, pages 6-1 through 6-22, Applied Biosystems, Part No. 901237; Ojwang, *et al.*, 1997, Biochemistry, 36:6033-6045). The phosphoramidite method of polynucleotide synthesis is the preferred method because of its efficient and rapid coupling and the stability of the starting materials. The synthesis is performed with the growing polynucleotide chain attached to a solid support, such that excess reagents, which are generally in the liquid phase, can be easily removed by washing, decanting, and/or filtration, thereby eliminating the need for purification steps between synthesis cycles.

The following briefly describes illustrative steps of a typical polynucleotide synthesis cycle using the phosphoramidite method. First, a solid support to which is attached a protected nucleoside monomer at its 3' terminus is treated with acid, *e.g.*, trichloroacetic acid, to remove the 5'-hydroxyl protecting group, freeing the hydroxyl group for a subsequent coupling reaction. After the coupling reaction is completed an activated intermediate is formed by contacting the support-bound nucleoside with a protected nucleoside phosphoramidite monomer and a weak acid, *e.g.*, tetrazole. The weak acid protonates the nitrogen atom of the phosphoramidite forming a reactive intermediate.

Nucleoside addition is generally complete within 30 seconds. Next, a capping step is performed, which terminates any polynucleotide chains that did not undergo nucleoside addition. Capping is preferably performed using acetic anhydride and 1-methylimidazole. The phosphite group of the internucleotide linkage is then converted to the more stable phosphotriester by oxidation using iodine as the preferred oxidizing agent and water as the oxygen donor. After oxidation, the hydroxyl protecting group of the newly added nucleoside is removed with a protic acid, *e.g.*, trichloroacetic acid or dichloroacetic acid, and the cycle is repeated one or more times until chain elongation is complete. After synthesis, the polynucleotide chain is cleaved from the support using a base, *e.g.*, ammonium hydroxide or *t*-butyl amine. The cleavage reaction also removes any phosphate protecting groups, *e.g.*, cyanoethyl. Finally, the protecting groups on the exocyclic amines of the bases and any protecting groups on the dyes are removed by treating the polynucleotide solution in base at an elevated temperature, *e.g.*, at about 55°C. Preferably the various protecting groups are removed using ammonium hydroxide or *t*-butyl amine.

Any of the nucleoside phosphoramidite monomers can be labeled using standard phosphoramidite chemistry methods (Hwang *et al.*, 1999, Proc. Natl. Acad. Sci. USA 96(23):12997-13002; Ojwang *et al.*, 1997, Biochemistry. 36:6033-6045 and references cited therein). Dye molecules useful for covalently coupling to phosphoramidites preferably comprise a primary hydroxyl group that is not part of the dye's chromophore. Illustrative dye molecules include, but are not limited to, disperse dye CAS 4439-31-0, disperse dye CAS 6054-58-6, disperse dye CAS 4392-69-2 (Sigma-Aldrich, St. Louis, MO), disperse red, and 1-pyrenebutanol (Molecular Probes, Eugene, OR). Other dyes useful for coupling to phosphoramidites will be apparent to those of skill in the art, such as fluorescein, cy3, and cy5 fluorescent dyes, and may be purchased from, *e.g.*, Sigma-Aldrich, St. Louis, MO or Molecular Probes, Inc., Eugene, OR.

In another embodiment, dye-labeled target RNA molecules are synthesized enzymatically using *in vitro* transcription (Hwang *et al.*, 1999, Proc. Natl. Acad. Sci. USA 96(23):12997-13002 and references cited therein). In this embodiment, a template DNA is denatured by heating to about 90°C and an oligonucleotide primer is annealed to the template DNA, for example by slow-cooling the mixture of the denatured template and the primer from about 90°C to room temperature. A mixture of ribonucleoside-5'-triphosphates capable of supporting template-directed enzymatic extension of the primed template (*e.g.*, a mixture including GTP, ATP, CTP, and UTP), including one or more dye-labeled ribonucleotides (Sigma-Aldrich, St. Louis, MO), is added to the primed template. Next, a polymerase enzyme is added to the mixture under conditions where the polymerase enzyme

is active, which are well-known to those skilled in the art. A labeled polynucleotide is formed by the incorporation of the labeled ribonucleotides during polymerase-mediated strand synthesis.

- 5 In yet another embodiment of the invention, nucleic acid molecules are end-labeled after their synthesis. Methods for labeling the 5'-end of an oligonucleotide include but are by no means limited to: (i) periodate oxidation of a 5'-to-5'-coupled ribonucleotide, followed by reaction with an amine-reactive label (Heller & Morisson, 1985, in *Rapid Detection and Identification of Infectious Agents*, D.T. Kingsbury and S. Falkow, eds., pp. 245-256, Academic Press); (ii) condensation of ethylenediamine with 5'-phosphorylated
10 polynucleotide, followed by reaction with an amine reactive label (Morrison, European Patent Application 232 967); (iii) introduction of an aliphatic amine substituent using an aminoethyl phosphite reagent in solid-phase DNA synthesis, followed by reaction with an amine reactive label (Cardullo *et al.*, 1988, Proc. Natl. Acad. Sci. USA 85:8790-8794); and
15 (iv) introduction of a thiophosphate group on the 5'-end of the nucleic acid, using phosphatase treatment followed by end-labeling with ATP- γ S and kinase, which reacts specifically and efficiently with maleimide-labeled fluorescent dyes (Czworkowski *et al.*, 1991, Biochem. 30:4821-4830).

- A detectable label should not be incorporated into a target nucleic acid at the
20 specific binding site at which test compounds are likely to bind, since the presence of a covalently attached label might interfere sterically or chemically with the binding of the test compounds at this site. Accordingly, if the region of the target nucleic acid that binds to a host cell factor is known, a detectable label is preferably incorporated into the nucleic acid molecule at one or more positions that are spatially or sequentially remote from the binding
25 region.

- After synthesis, the labeled target nucleic acid can be purified using standard techniques known to those skilled in the art (*see* Hwang *et al.*, 1999, Proc. Natl. Acad. Sci. USA 96(23):12997-13002 and references cited therein). Depending on the length of the target nucleic acid and the method of its synthesis, such purification techniques include, but
30 are not limited to, reverse-phase high-performance liquid chromatography ("reverse-phase HPLC"), fast performance liquid chromatography ("FPLC"), and gel purification. After purification, the target RNA is refolded into its native conformation, preferably by heating to approximately 85-95°C and slowly cooling to room temperature in a buffer, *e.g.*, a buffer comprising about 50 mM Tris-HCl, pH 8 and 100 mM NaCl.

- 35 In another embodiment, the target nucleic acid can also be radiolabeled. A radiolabel, such as, but not limited to, an isotope of phosphorus, sulfur, or hydrogen, may be

incorporated into a nucleotide, which is added either after or during the synthesis of the target nucleic acid. Methods for the synthesis and purification of radiolabeled nucleic acids are well known to one of skill in the art. See, e.g., Sambrook *et al.*, 1989, in *Molecular Cloning: A Laboratory Manual*, pp 10.2-10.70, Cold Spring Harbor Laboratory Press, and the references cited therein, which are hereby incorporated by reference in their entireties.

In another embodiment, the target nucleic acid can be attached to an inorganic nanoparticle. A nanoparticle is a cluster of ions with controlled size from 0.1 to 1000 nm comprised of metals, metal oxides, or semiconductors including, but not limited to Ag_2S , ZnS , CdS , CdTe , Au , or TiO_2 . Nanoparticles have unique optical, electronic and catalytic properties relative to bulk materials which can be adjusted according to the size of the particle. Methods for the attachment of nucleic acids are well known to one of skill in the art (see, e.g., Niemeyer, 2001, *Angew. Chem. Int. Ed.* 40: 4129-4158, International Patent Publication WO/0218643, and the references cited therein, the disclosures of which are hereby incorporated by reference in their entireties).

5.3. Libraries of Small Molecules

Libraries screened using the methods of the present invention can comprise a variety of types of test compounds. In some embodiments, the test compounds are nucleic acid or peptide molecules. In a non-limiting example, peptide molecules can exist in a phage display library. In other embodiments, types of test compounds include, but are not limited to, peptide analogs including peptides comprising non-naturally occurring amino acids, e.g., D-amino acids, phosphorous analogs of amino acids, such as α -amino phosphoric acids and α -amino phosphonic acids, or amino acids having non-peptide linkages, nucleic acid analogs such as phosphorothioates and PNAs, hormones, antigens, synthetic or naturally occurring drugs, opiates, dopamine, serotonin, catecholamines, thrombin, acetylcholine, prostaglandins, organic molecules, pheromones, adenosine, sucrose, glucose, lactose and galactose. Libraries of polypeptides or proteins can also be used.

In a preferred embodiment, the combinatorial libraries are small organic molecule libraries, such as, but not limited to, benzodiazepines, isoprenoids, thiazolidinones, metathiazanones, pyrrolidines, morpholino compounds, and diazepindiones. In another embodiment, the combinatorial libraries comprise peptoids; random bio-oligomers; diversomers such as hydantoins, benzodiazepines and dipeptides; vinylous polypeptides; nonpeptidal peptidomimetics; oligocarbamates; peptidyl phosphonates; peptide nucleic acid libraries; antibody libraries; or carbohydrate libraries.

Combinatorial libraries are themselves commercially available (see, e.g., Advanced ChemTech Europe Ltd., Cambridgeshire, UK; ASINEX, Moscow Russia; BioFocus plc, Sittingbourne, UK; Bionet Research (A division of Key Organics Limited), Camelford, UK; ChemBridge Corporation, San Diego, California; ChemDiv Inc, San Diego, California; ChemRx Advanced Technologies, South San Francisco, California; ComGenex Inc., Budapest, Hungary; Evotec OAI Ltd, Abingdon, UK; IF LAB Ltd., Kiev, Ukraine; Maybridge plc, Cornwall, UK; PharmaCore, Inc., North Carolina; SDDCO Inc, Tucson, Arizona; TimTec Inc, Newark, Delaware; Tripos Receptor Research Ltd, Bude, UK; Toslab, Ekaterinburg, Russia).

In one embodiment, the combinatorial compound library for the methods of the present invention may be synthesized. There is a great interest in synthetic methods directed toward the creation of large collections of small organic compounds, or libraries, which could be screened for pharmacological, biological or other activity (Dolle, 2001, J. Comb. Chem. 3:477-517; Hall *et al.*, 2001, J. Comb. Chem. 3:125-150; Dolle, 2000, J. Comb. Chem. 2:383-433; Dolle, 1999, J. Comb. Chem. 1:235-282). The synthetic methods applied to create vast combinatorial libraries are performed in solution or in the solid phase, *i.e.*, on a solid support. Solid-phase synthesis makes it easier to conduct multi-step reactions and to drive reactions to completion with high yields because excess reagents can be easily added and washed away after each reaction step. Solid-phase combinatorial synthesis also tends to improve isolation, purification and screening. However, the more traditional solution phase chemistry supports a wider variety of organic reactions than solid-phase chemistry. Methods and strategies for the synthesis of combinatorial libraries can be found in *A Practical Guide to Combinatorial Chemistry*, A.W. Czarnik and S.H. Dewitt, eds., American Chemical Society, 1997; *The Combinatorial Index*, B.A. Bunin, Academic Press, 1998; *Organic Synthesis on Solid Phase*, F.Z. Dörwald, Wiley-VCH, 2000; and *Solid-Phase Organic Syntheses, Vol. 1*, A.W. Czarnik, ed., Wiley Interscience, 2001.

Combinatorial compound libraries of the present invention may be synthesized using apparatuses described in US Patent No. 6,358,479 to Frisina *et al.*, U.S. Patent No. 6,190,619 to Kilcoin *et al.*, US Patent No. 6,132,686 to Gallup *et al.*, US Patent No. 6,126,904 to Zuellig *et al.*, US Patent No. 6,074,613 to Harness *et al.*, US Patent No. 6,054,100 to Stanchfield *et al.*, and US Patent No. 5,746,982 to Saneii *et al.* which are hereby incorporated by reference in their entirety. These patents describe synthesis apparatuses capable of holding a plurality of reaction vessels for parallel synthesis of multiple discrete compounds or for combinatorial libraries of compounds.

In one embodiment, the combinatorial compound library can be synthesized in solution. The method disclosed in U.S. Patent No. 6,194,612 to Boger *et al.*, which is hereby incorporated by reference in its entirety, features compounds useful as templates for solution phase synthesis of combinatorial libraries. The template is designed to permit reaction products to be easily purified from unreacted reactants using liquid/liquid or solid/liquid extractions. The compounds produced by combinatorial synthesis using the template will preferably be small organic molecules. Some compounds in the library may mimic the effects of non-peptides or peptides. In contrast to solid phase synthesis of combinatorial compound libraries, liquid phase synthesis does not require the use of specialized protocols for monitoring the individual steps of a multistep solid phase synthesis (Egner *et al.*, 1995, J. Org. Chem. 60:2652; Anderson *et al.*, 1995, J. Org. Chem. 60:2650; Fitch *et al.*, 1994, J. Org. Chem. 59:7955; Look *et al.*, 1994, J. Org. Chem. 49:7588; Metzger *et al.*, 1993, Angew. Chem., Int. Ed. Engl. 32:894; Youngquist *et al.*, 1994, Rapid Commun. Mass Spect. 8:77; Chu *et al.*, 1995, J. Am. Chem. Soc. 117:5419; Brummel *et al.*, 1994, Science 264:399; Stevanovic *et al.*, 1993, Bioorg. Med. Chem. Lett. 3:431).

Combinatorial compound libraries useful for the methods of the present invention can be synthesized on solid supports. In one embodiment, a split synthesis method, a protocol of separating and mixing solid supports during the synthesis, is used to synthesize a library of compounds on solid supports (*see* Lam *et al.*, 1997, Chem. Rev. 97:41-448; Ohlmeyer *et al.*, 1993, Proc. Natl. Acad. Sci. USA 90:10922-10926 and references cited therein). Each solid support in the final library has substantially one type of test compound attached to its surface. Other methods for synthesizing combinatorial libraries on solid supports, wherein one product is attached to each support, will be known to those of skill in the art (*see*, e.g., Nefzi *et al.*, 1997, Chem. Rev. 97:449-472 and US Patent No. 6,087,186 to Cargill *et al.* which are hereby incorporated by reference in their entirety).

As used herein, the term "solid support" is not limited to a specific type of solid support. Rather a large number of supports are available and are known to one skilled in the art. Solid supports include silica gels, resins, derivatized plastic films, glass beads, cotton, plastic beads, polystyrene beads, alumina gels, and polysaccharides. A suitable solid support may be selected on the basis of desired end use and suitability for various synthetic protocols. For example, for peptide synthesis, a solid support can be a resin such as p-methylbenzhydrylamine (pMBHA) resin (Peptides International, Louisville, KY), polystyrenes (e.g., PAM-resin obtained from Bachem Inc., Peninsula Laboratories, etc.), including chloromethylpolystyrene, hydroxymethylpolystyrene and

aminomethylpolystyrene, poly (dimethylacrylamide)-grafted styrene co-divinyl-benzene (e.g., POLYHIPE resin, obtained from Aminotech, Canada), polyamide resin (obtained from Peninsula Laboratories), polystyrene resin grafted with polyethylene glycol (e.g., TENTAGEL or ARGOGEL, Bayer, Tübingen, Germany) polydimethylacrylamide resin (obtained from Milligen/Bioscience, California), or Sepharose (Pharmacia, Sweden).

In one embodiment, the solid phase support is suitable for *in vivo* use, i.e., it can serve as a carrier or support for administration of the test compound to a patient (e.g., TENTAGEL, Bayer, Tübingen, Germany). In a particular embodiment, the solid support is palatable and/or orally ingestible.

In some embodiments of the present invention, compounds can be attached to solid supports via linkers. Linkers can be integral and part of the solid support, or they may be nonintegral that are either synthesized on the solid support or attached thereto after synthesis. Linkers are useful not only for providing points of test compound attachment to the solid support, but also for allowing different groups of molecules to be cleaved from the solid support under different conditions, depending on the nature of the linker. For example, linkers can be, *inter alia*, electrophilically cleaved, nucleophilically cleaved, photocleavable, enzymatically cleaved, cleaved by metals, cleaved under reductive conditions or cleaved under oxidative conditions.

In another embodiment, the combinatorial compound libraries can be assembled *in situ* using dynamic combinatorial chemistry as described in European Patent Application 1,118,359 A1 to Lehn; Huc & Nguyen, 2001, Comb. Chem. High Throughput. Screen. 4:53-74; Lehn and Eliseev, 2001, Science 291:2331-2332; Cousins *et al.* 2000, Curr. Opin. Chem. Biol. 4: 270-279; and Karan & Miller, 2000, Drug. Disc. Today 5:67-75 which are incorporated by reference in their entirety.

Dynamic combinatorial chemistry uses non-covalent interaction with a target biomolecule, including but not limited to a protein, RNA, or DNA, to favor assembly of the most tightly binding molecule that is a combination of constituent subunits present as a mixture in the presence of the biomolecule. According to the laws of thermodynamics, when a collection of molecules is able to combine and recombine at equilibrium through reversible chemical reactions in solution, molecules, preferably one molecule, that bind most tightly to a templating biomolecule will be present in greater amount than all other possible combinations. The reversible chemical reactions include, but are not limited to, imine, acyl-hydrazone, amide, acetal, or ester formation between carbonyl-containing compounds and amines, hydrazines, or alcohols; thiol exchange between disulfides; alcohol

exchange in borate esters; Diels-Alder reactions; thermal- or photoinduced sigmatropic or electrocyclic rearrangements; or Michael reactions.

In the preferred embodiment of this technique, the constituent components of the dynamic combinatorial compound library are allowed to combine and reach equilibrium in the absence of the target RNA and then incubated in the presence of the target RNA, preferably at physiological conditions, until a second equilibrium is reached. The second, perturbed, equilibrium (the so-called "templated mixture") can, but need not necessarily, be fixed by a further chemical transformation, including but not limited to reduction, oxidation, hydrolysis, acidification, or basification, to prevent restoration of the original equilibrium when the dynamical combinatorial compound library is separated from the target RNA.

In the preferred embodiment of this technique, the predominant product or products of the templated dynamic combinatorial library can be separated from the minor products and directly identified. In another embodiment, the identity of the predominant product or products can be identified by a deconvolution strategy involving preparation of derivative dynamic combinatorial libraries, as described in European Patent Application 1,118,359 A1, which is incorporated by reference in its entirety, whereby each component of the mixture is, preferably one-by-one but possibly group-wise, left out of the mixture and the ability of the derivative library mixture at chemical equilibrium to bind the target RNA is measured. The components whose removal most greatly reduces the ability of the derivative dynamic combinatorial library to bind the target RNA are likely the components of the predominant product or products in the original dynamic combinatorial library.

25

5.4. Library Screening

After a target nucleic acid, such as but not limited to RNA or DNA, is labeled and a test compound library is synthesized or purchased or both, the labeled target nucleic acid is used to screen the library to identify test compounds that bind to the nucleic acid. Screening comprises contacting a labeled target nucleic acid with an individual, or small group, of the components of the compound library. Preferably, the contacting occurs in an aqueous solution, and most preferably, under physiologic conditions. The aqueous solution preferably stabilizes the labeled target nucleic acid and prevents denaturation or degradation of the nucleic acid without interfering with binding of the test compounds. The aqueous solution can be similar to the solution in which a complex between the target RNA and its corresponding host cell factor (if known) is formed *in vitro*. For example, TK

35

buffer, which is commonly used to form Tat protein-TAR RNA complexes *in vitro*, can be used in the methods of the invention as an aqueous solution to screen a library of test compounds for TAR RNA binding compounds.

5 The methods of the present invention for screening a library of test compounds preferably comprise contacting a test compound with a target nucleic acid in the presence of an aqueous solution, the aqueous solution comprising a buffer and a combination of salts, preferably approximating or mimicking physiologic conditions. The aqueous solution optionally further comprises non-specific nucleic acids, such as, but not
10 limited to, DNA; yeast tRNA; salmon sperm DNA; homoribopolymers such as, but not limited to, poly IC, polyA, polyU, and polyC; and non-specific RNA. The non-specific RNA may be an unlabeled target nucleic acid having a mutation at the binding site, which renders the unlabeled nucleic acid incapable of interacting with a test compound at that site. For example, if dye-labeled TAR RNA is used to screen a library, unlabeled TAR RNA
15 having a mutation in the uracil 23/cytosine 24 bulge region may also be present in the aqueous solution. Without being bound by any theory, the addition of unlabeled RNA that is essentially identical to the dye-labeled target RNA except for a mutation at the binding site might minimize interactions of other regions of the dye-labeled target RNA with test compounds or with the solid support and prevent false positive results.

20 The solution further comprises a buffer, a combination of salts, and optionally, a detergent or a surfactant. The pH of the solution typically ranges from about 5 to about 8, preferably from about 6 to about 8, most preferably from about 6.5 to about 8. A variety of buffers may be used to achieve the desired pH. Suitable buffers include, but are not limited to, Tris, Mes, Bis-Tris, Ada, Aces, Pipes, Mopso, Bis-Tris propane, Bes,
25 Mops, Tes, Hepes, Dipso, Mobs, Tapso, Trizma, Heppso, Popso, TEA, Epps, Tricine, Gly-Gly, Bicine, and sodium-potassium phosphate. The buffering agent comprises from about 10 mM to about 100 mM, preferably from about 25 mM to about 75 mM, most preferably from about 40 mM to about 60 mM buffering agent. The pH of the aqueous solution can be optimized for different screening reactions, depending on the target RNA used and the
30 types of test compounds in the library, and therefore, the type and amount of the buffer used in the solution can vary from screen to screen. In a preferred embodiment, the aqueous solution has a pH of about 7.4, which can be achieved using about 50 mM Tris buffer.

35 In addition to an appropriate buffer, the aqueous solution further comprises a combination of salts, from about 0 mM to about 100 mM KCl, from about 0 mM to about 1 M NaCl, and from about 0 mM to about 200 mM MgCl₂. In a preferred embodiment, the combination of salts is about 100 mM KCl, 500 mM NaCl, and 10 mM MgCl₂. Without

being bound by any theory, Applicant has found that a combination of KCl, NaCl, and MgCl₂ stabilizes the target RNA such that most of the RNA is not denatured or digested over the course of the screening reaction. The optional concentration of each salt used in the aqueous solution is dependent on the particular target RNA used and can be determined using routine experimentation.

The solution optionally comprises from about 0.01% to about 0.5% (w/v) of a detergent or a surfactant. Without being bound by any theory, a small amount of detergent or surfactant in the solution might reduce non-specific binding of the target RNA to the solid support and control aggregation and increase stability of target RNA molecules. Typical detergents useful in the methods of the present invention include, but are not limited to, anionic detergents, such as salts of deoxycholic acid, 1-heptanesulfonic acid, N-laurylsarcosine, lauryl sulfate, 1-octane sulfonic acid and taurocholic acid; cationic detergents such as benzalkonium chloride, cetylpyridinium, methylbenzethonium chloride, and decamethonium bromide; zwitterionic detergents such as CHAPS, CHAPSO, alkyl betaines, alkyl amidoalkyl betaines, N-dodecyl-N,N-dimethyl-3-ammonio-1-propanesulfonate, and phosphatidylcholine; and non-ionic detergents such as n-decyl α-D-glucopyranoside, n-decyl β-D-maltopyranoside, n-dodecyl β-D-maltoside, n-octyl β-D-glucopyranoside, sorbitan esters, n-tetradecyl β-D-maltoside, octylphenoxy polyethoxyethanol (Nonidet P-40), nonylphenoxypolyethoxyethanol (NP-40), and tritons. Preferably, the detergent, if present, is a nonionic detergent. Typical surfactants useful in the methods of the present invention include, but are not limited to, ammonium lauryl sulfate, polyethylene glycols, butyl glucoside, decyl glucoside, Polysorbate 80, lauric acid, myristic acid, palmitic acid, potassium palmitate, undecanoic acid, lauryl betaine, and lauryl alcohol. More preferably, the detergent, if present, is Triton X-100 and present in an amount of about 0.1% (w/v).

Non-specific binding of a labeled target nucleic acid to test compounds can be further minimized by treating the binding reaction with one or more blocking agents. In one embodiment, the binding reactions are treated with a blocking agent, e.g., bovine serum albumin ("BSA"), before contacting with the labeled target nucleic acid. In another embodiment, the binding reactions are treated sequentially with at least two different blocking agents. This blocking step is preferably performed at room temperature for from about 0.5 to about 3 hours. In a subsequent step, the reaction mixture is further treated with unlabeled RNA having a mutation at the binding site. This blocking step is preferably performed at about 4°C for from about 12 hours to about 36 hours before addition of the dye-labeled target RNA. Preferably, the solution used in the one or more blocking steps is

substantially similar to the aqueous solution used to screen the library with the dye-labeled target RNA, *e.g.*, in pH and salt concentration.

Once contacted, the mixture of labeled target nucleic acid and the test compound is preferably maintained at 4°C for from about 1 day to about 5 days, preferably from about 2 days to about 3 days with constant agitation. To identify the reactions in which binding to the labeled target nucleic acid occurred, after the incubation period, bound from free compounds are determined using an electrophoretic technique (see Section 5.5.1), or any of the methods disclosed in Section 5.5 *infra*. In another embodiment, the complexed target nucleic acid does not need to be separated from the free target nucleic acid if a technique (*i.e.*, spectrometry) that differentiates between bound and unbound target nucleic acids is used.

The methods for identifying small molecules bound to labeled nucleic acid will vary with the type of label on the target nucleic acid. For example, if a target RNA is labeled with a visible or fluorescent dye, the target RNA complexes are preferably identified using a chromatographic technique that separates bound from free target by an electrophoretic or size differential technique using individual reactions. The reactions corresponding to changes in the migration of the complexed RNA can be cross-referenced to the small molecule compound(s) added to said reaction. Alternatively, complexed target RNA can be screened *en masse* and then separated from free target RNA using an electrophoretic or size differential technique, the resultant complexed target is then analyzed using a mass spectrometric technique. In this fashion the bound small molecule can be identified on the basis of its molecular weight. In this reaction *a priori* knowledge of the exact molecular weights of all compounds within the library is known. In another embodiment, the test compounds bound to the target nucleic acid may not require separation from the unbound target nucleic acid if a technique such as, but not limited to, spectrometry is used.

5.5. Separation Methods for Screening Test Compounds

Any method that detects an altered physical property of a target nucleic acid complexed to a test compound from the unbound target nucleic acid may be used for separation of the complexed and non-complexed target nucleic acids. Methods that can be utilized for the physical separation of complexed target RNA from unbound target RNA include, but are not limited to, electrophoresis, fluorescence spectroscopy, surface plasmon resonance, mass spectrometry, scintillation, proximity assay, structure-activity relationships ("SAR") by NMR spectroscopy, size exclusion chromatography, affinity chromatography,

and nanoparticle aggregation.

5.5.1. Electrophoresis

5 Methods for separation of the complex of a target RNA bound to a test
compound from the unbound RNA comprises any method of electrophoretic separation,
including but not limited to, denaturing and non-denaturing polyacrylamide gel
electrophoresis, urea gel electrophoresis, gel filtration, pulsed field gel electrophoresis, two
dimensional gel electrophoresis, continuous flow electrophoresis, zone electrophoresis,
10 agarose gel electrophoresis, and capillary electrophoresis.

 In a preferred embodiment, an automated electrophoretic system comprising
a capillary cartridge having a plurality of capillary tubes is used for high-throughput
screening of test compounds bound to target RNA. Such an apparatus for performing
automated capillary gel electrophoresis is disclosed in U.S. Patent Nos. 5,885,430;
15 5,916,428; 6,027,627; and 6,063,251, the disclosures of which are incorporated by
reference in their entireties.

 The device disclosed in U.S. Patent No. 5,885,430, which is incorporated by
reference in its entirety, allows one to simultaneously introduce samples into a plurality of
capillary tubes directly from microtiter trays having a standard size. U.S. Patent No.
20 5,885,430 discloses a disposable capillary cartridge which can be cleaned between
electrophoresis runs, the cartridge having a plurality of capillary tubes. A first end of each
capillary tube is retained in a mounting plate, the first ends collectively forming an array in
the mounting plate. The spacing between the first ends corresponds to the spacing between
the centers of the wells of a microtiter tray having a standard size. Thus, the first ends of
25 the capillary tubes can simultaneously be dipped into the samples present in the tray's wells.
The cartridge is provided with a second mounting plate in which the second ends of the
capillary tubes are retained. The second ends of the capillary tubes are arranged in an array
which corresponds to the wells in the microtiter tray, which allows for each capillary tube
to be isolated from its neighbors and therefore free from cross-contamination, as each end
30 is dipped into an individual well.

 Plate holes may be provided in each mounting plate and the capillary tubes
inserted through these plate holes. In such a case, the plate holes are sealed airtight so that
the side of the mounting plate having the exposed capillary ends can be pressurized.
Application of a positive pressure in the vicinity of the capillary openings in this mounting
35 plate allows for the introduction of air and fluids during electrophoretic operations and also
can be used to force out gel and other materials from the capillary tubes during

reconditioning. The capillary tubes may be protected from damage using a needle comprising a cannula and/or plastic tubes, and the like when they are placed in these plate holes. When metallic cannula or the like are used, they can serve as electrical contacts for current flow during electrophoresis. In the presence of a second mounting plate, the second mounting plate is provided with plate holes through which the second ends of the capillary tubes project. In this instance, the second mounting plate serves as a pressure containment member of a pressure cell and the second ends of the capillary tubes communicate with an internal cavity of the pressure cell. The pressure cell is also formed with an inlet and an outlet. Gels, buffer solutions, cleaning agents, and the like may be introduced into the internal cavity through the inlet, and each of these can simultaneously enter the second ends of the capillaries.

In another preferred embodiment, the automated electrophoretic system can comprise a chip system consisting of complex designs of interconnected channels that perform and analyze enzyme reactions using part of a channel design as a tiny, continuously operating electrophoresis material, where reactions with one sample are going on in one area of the chip while electrophoretic separation of the products of another sample is taking place in a different part of the chip. Such a system is disclosed in U.S. Patent Nos. 5,699,157; 5,842,787; 5,869,004; 5,876,675; 5,942,443; 5,948,227; 6,042,709; 6,042,710; 6,046,056; 6,048,498; 6,086,740; 6,132,685; 6,150,119; 6,150,180; 6,153,073; 6,167,910; 6,171,850; and 6,186,660, the disclosures of which are incorporated by reference in their entireties.

The system disclosed in U.S. Patent No. 5,699,157, which is hereby incorporated by reference in its entirety, provides for a microfluidic system for high-speed electrophoretic analysis of subject materials for applications in the fields of chemistry, biochemistry, biotechnology, molecular biology and numerous other areas. The system has a channel in a substrate, a light source and a photoreceptor. The channel holds subject materials in solution in an electric field so that the materials move through the channel and separate into bands according to species. The light source excites fluorescent light in the species bands and the photoreceptor is arranged to receive the fluorescent light from the bands. The system further has a means for masking the channel so that the photoreceptor can receive the fluorescent light only at periodically spaced regions along the channel. The system also has an unit connected to analyze the modulation frequencies of light intensity received by the photoreceptor so that velocities of the bands along the channel are determined, which allows the materials to be analyzed.

The system disclosed in U.S. Patent No. 5,699,157 also provides for a

method of performing high-speed electrophoretic analysis of subject materials, which comprises the steps of holding the subject materials in solution in a channel of a microfluidic system; subjecting the materials to an electric field so that the subject materials move through the channel and separate into species bands; directing light toward the channel; receiving light from periodically spaced regions along the channel simultaneously; and analyzing the frequencies of light intensity of the received light so that velocities of the bands along the channel can be determined for analysis of said materials. The determination of the velocity of a species band determines the electrophoretic mobility of the species and its identification.

U.S. Patent No. 5,842,787, which is hereby incorporated by reference in its entirety, is generally directed to devices and systems employ channels having, at least in part, depths that are varied over those which have been previously described (such as the device disclosed in U.S. Patent No. 5,699,157), wherein said channel depths provide numerous beneficial and unexpected results such as but not limited to, a reduction in sample perturbation, reduced non-specific sample mixture by diffusion, and increased resolution.

In another embodiment, the electrophoretic method of separation comprises polyacrylamide gel electrophoresis. In a preferred embodiment, the polyacrylamide gel electrophoresis is non-denaturing, so as to differentiate the mobilities of the target RNA bound to a test compound from free target RNA. If the polyacrylamide gel electrophoresis is denaturing, then the target RNA:test compound complex must be cross-linked prior to electrophoresis to prevent the disassociation of the target RNA from the test compound during electrophoresis. Such techniques are well known to one of skill in the art.

In one embodiment of the method, the binding of test compounds to target nucleic acid can be detected, preferably in an automated fashion, by gel electrophoretic analysis of interference footprinting. RNA can be degraded at specific base sites by enzymatic methods such as ribonucleases A, U₂, CL₃, T₁, Phy M, and *B. cereus* or chemical methods such as diethylpyrocarbonate, sodium hydroxide, hydrazine, piperidine formate, dimethyl sulfate, [2,12-dimethyl-3,7,11,17-tetraazacyclo[11.3.1]heptadeca-1(17),2,11,13,15-pentaenato] nickel(II) (NiCR), cobalt(II)chloride, or iron(II) ethylenediaminetetraacetate (Fe-EDTA) as described for example in Zheng *et al.*, 1999, *Biochem.* 37:2207-2214; Latham & Cech, 1989, *Science* 245:276-282; and Sambrook *et al.*, 2001, in *Molecular Cloning: A Laboratory Manual*, pp 12.61-12.73, Cold Spring Harbor Laboratory Press, and the references cited therein, which are hereby incorporated by reference in their entirety. The

specific pattern of cleavage sites is determined by the accessibility of particular bases to the reagent employed to initiate cleavage and, as such, is therefore determined by the three-dimensional structure of the RNA.

5 The interaction of small molecules with a target nucleic acid can change the accessibility of bases to these cleavage reagents both by causing conformational changes in the target nucleic acid or by covering a base at the binding interface. When a test compound binds to the nucleic acid and changes the accessibility of bases to cleavage reagents, the observed cleavage pattern will change. This method can be used to identify and characterize the binding of small molecules to RNA as described, for example, by
10 Prudent *et al.*, 1995, J. Am. Chem. Soc. 117:10145-10146 and Mei *et al.*, 1998, Biochem. 37:14204-14212.

 In the preferred embodiment of this technique, the detectably labeled target nucleic acid is incubated with an individual test compound and then subjected to treatment with a cleavage reagent, either enzymatic or chemical. The reaction mixture can be
15 preferably be examined directly, or treated further to isolate and concentrate the nucleic acid. The fragments produced are separated by electrophoresis and the pattern of cleavage can be compared to a cleavage reaction performed in the absence of test compound. A change in the cleavage pattern directly indicates that the test compound binds to the target
20 nucleic acid. Multiple test compounds can be examined both in parallel and serially.

 Other embodiments of electrophoretic separation include, but are not limited to urea gel electrophoresis, gel filtration, pulsed field gel electrophoresis, two dimensional gel electrophoresis, continuous flow electrophoresis, zone electrophoresis, and agarose gel electrophoresis.

25

5.5.2. Fluorescence Spectroscopy

 In a preferred embodiment, fluorescence polarization spectroscopy, an optical detection method that can differentiate the proportion of a fluorescent molecule that is either bound or unbound in solution (e.g., the labeled target nucleic acid of the present
30 invention), can be used to read reaction results without electrophoretic separation of the samples. Fluorescence polarization spectroscopy can be used to read the reaction results in the chip system disclosed in U.S. Patent Nos. 5,699,157; 5,842,787; 5,869,004; 5,876,675; 5,942,443; 5,948,227; 6,042,709; 6,042,710; 6,046,056; 6,048,498; 6,086,740; 6,132,685; 6,150,119; 6,150,180; 6,153,073; 6,167,910; 6,171,850; and 6,186,660, the disclosures of
35 which are incorporated by reference in their entireties. The application of fluorescence

polarization spectroscopy to the chip system disclosed in the U.S. Patents listed *supra* is fast, efficient, and well-adapted for high-throughput screening.

5 In another embodiment, a compound that has an affinity for the target nucleic acid of interest can be labeled with a fluorophore to screen for test compounds that bind to the target nucleic acid. For example, a pyrene-containing aminoglycoside analog was used to accurately monitor antagonist binding to a prokaryotic 16S rRNA A site (which comprises the natural target for aminoglycoside antibiotics) in a screen using a fluorescence quenching technique in a 96-well plate format (Hamasaki & Rando, 1998, Anal. Biochem. 261(2):183-90).

10 In another embodiment, fluorescence resonance energy transfer (FRET) can be used to screen for test compounds that bind to the target nucleic acid. FRET, a characteristic change in fluorescence, occurs when two fluorophores with overlapping emission and excitation wavelength bands are held together in close proximity, such as by a binding event. In the preferred embodiment, the fluorophore on the target nucleic acid and the fluorophore on the test compounds will have overlapping excitation and emission spectra such that one fluorophore (the donor) transfers its emission energy to excite the other fluorophore (the acceptor). The acceptor preferably emits light of a different wavelength upon relaxing to the ground state, or relaxes non-radiatively to quench fluorescence. FRET is very sensitive to the distance between the two fluorophores, and allows measurement of molecular distances less than 10 nm. For example, U.S. Patent 6,337,183 to Arenas *et al.*, which is incorporated by reference in its entirety, describes a screen for compounds that bind RNA that uses FRET to measure the effect of test compounds on the stability of a target RNA molecule where the target RNA is labeled with both fluorescent acceptor and donor molecules and the distance between the two fluorophores as determined by FRET provides a measure of the folded structure of the RNA. Matsumoto *et al.* (2000, Bioorg. Med. Chem. Lett. 10:1857-1861) describe a system where a peptide that binds to HIV-1 TAR RNA is labeled on one end with a fluorescein fluorophore and a tetramethylrhodamine on the other end. The conformational change of the peptide upon binding to the RNA provided a FRET signal to screen for compounds that bound to the TAR RNA.

25 In the preferred embodiment, both the target nucleic acid and a compound that has an affinity for the target nucleic acid of interest are labeled with fluorophores with overlapping emission and excitation spectra (donor and acceptor), including but not limited to fluorescein and derivatives, rhodamine and derivatives, cyanine dyes and derivatives, bora-3a,4a-diaza-s-indacene (BODIPY®) and derivatives, pyrene, nanoparticles, or

non-fluorescent quenching molecules. Binding of a labeled test compound to the target nucleic acid can be identified by the change in observable fluorescence as a result of FRET.

If the target nucleic acid is labeled with the donor fluorophore, then the test compounds is labeled with the acceptor fluorophore. Conversely, if the target nucleic acid is labeled with the acceptor fluorophore, then the test compounds is labeled with the donor fluorophore. A wide variety of labels may be used, with the choice of label depending on sensitivity required, ease of conjugation with the compound, stability requirements, available instrumentation, and disposal provisions. The fluorophore on the target nucleic acid must be in close proximity to the binding site of the test compounds, but should not be incorporated into a target nucleic acid at the specific binding site at which test compounds are likely to bind, since the presence of a covalently attached label might interfere sterically or chemically with the binding of the test compounds at this site.

In yet another embodiment, homogeneous time-resolved fluorescence ("HTRF") techniques based on time-resolved energy transfer from lanthanide ion complexes to a suitable acceptor species can be adapted for high-throughput screening for inhibitors of RNA-protein complexes (Hemmilä, 1999, J. Biomol. Screening 4:303-307; Mathis, 1999, J. Biomol. Screening 4:309-313). HTRF is similar to fluorescence resonance energy transfer using conventional organic dye pairs, but has several advantages, such as increased sensitivity and efficiency, and background elimination (Xavier *et al.*, 2000, Trends Biotechnol. 18(8):349-356).

Fluorescence spectroscopy has traditionally been used to characterize DNA-protein and protein-protein interactions, but fluorescence spectroscopy has not been widely used to characterize RNA-protein interactions because of an interfering absorption of RNA nucleotides with the intrinsic tryptophan fluorescence of proteins (Xavier *et al.*, 2000, Trends Biotechnol. 18(8):349-356.). However, fluorescence spectroscopy has been used in studying the single tryptophan residue within the arginine-rich RNA-binding domain of Rev protein and its interaction with the RRE in a time-resolved fluorescence study (Kwon & Carson, 1998, Anal. Biochem. 264:133-140). Thus, in this invention, fluorescence spectroscopy is less preferred if the test compounds or peptides or proteins possess intrinsic tryptophan fluorescence. However, fluorescence spectroscopy can be used for test compounds that do not possess intrinsic fluorescence.

5.5.3. Surface Plasmon Resonance ("SPR")

Surface plasmon resonance (SPR) can be used for determining kinetic rate constants and equilibrium constants for macromolecular interactions by following the

association project in "real time" (Schuck, 1997, *Annu. Rev. Biophys. Biomol. Struct.* 26:541-566).

The principle of SPR is summarized by Xavier *et al.* (*Trends Biotechnol.*, 2000, 18(8):349-356) as follows. Total internal reflection occurs at the boundary between two substances of different refractive index. The incident light's electromagnetic field penetrates beyond the interface as an evanescent wave, which extends a few hundred nanometers beyond the surface into the medium. Insertion of a thin gold foil at the interface produced SPR owing to the absorption of the energy from the evanescent wave by free electron clouds of the metal (plasmons). As a result of this absorbance, there is a drop in the intensity of the reflected light at a particular angle of incidence. The evanescent wave profile depends exquisitely on the refractive index of the medium it probes. Thus, the angle at which absorption occurs is very sensitive to the refractive changes in the external medium. All proteins and nucleic acids are known to change the refractive index of water by a similar amount per unit mass, irrespective of their amino acid or nucleotide composition (the refractive index change is different for proteins and nucleic acids). When the protein or nucleic acid content of the layer at the sensor changes, the refractive index also changes. Typically, one member of a complex is immobilized in a dextran layer and then the other member is introduced into the solution, either in a flow cell (Biacore AB, Uppsala, Sweden) or a stirred cuvette (Affinity Sensors, Santa Fe, New Mexico). It has been determined that there is a linear correlation between the surface concentration of protein or nucleic acid and the shift in resonance angle, which can be used to quantitate kinetic rate constants and/or the equilibrium constants.

In the present invention, the target RNA may be immobilized to the sensor surface through a streptavidin-biotin linkage, the linkage of which is disclosed by Crouch *et al.* (*Methods Mol. Biol.*, 1999, 118:143-160). The RNA is biotinylated either during synthesis or post-synthetically via the conversion of the 3' terminal ribonucleoside of the RNA into a reactive free amino group or using a T7 polymerase incorporated guanosine monophosphorothioate at the 5' end. SPR has been used to determine the stoichiometry and affinity of the interaction between the HIV Rev protein and the RRE (Van Ryk & Venkatesan, 1999, *J. Biol. Chem.* 274:17452-17463) and the aminoglycoside antibiotics with RRE and a model RNA derived from the 16S ribosomal A site, respectively (Hendrix *et al.*, 1997, *J. Am. Chem. Soc.* 119:3641-3648; Wong *et al.*, 1998, *Chem. Biol.* 5:397-406).

In one embodiment of the present invention, the target nucleic acid can be immobilized to a sensor surface (*e.g.*, by a streptavidin-biotin linkage) and SPR can be used

to (a) determine whether the target RNA binds a test compound and (b) further characterize the binding of the target nucleic acids of the present invention to a test compound.

5.5.4. Mass Spectrometry

An automated method for analyzing mass spectrometer data which can analyze complex mixtures containing many thousands of components and can correct for background noise, multiply charged peaks and atomic isotope peaks is described in U.S. Patent No. 6,147,344, which is hereby incorporated by reference in its entirety. The system disclosed in U.S. Patent No. 6,147,344 is a method for analyzing mass spectrometer data in which a control sample measurement is performed providing a background noise check. The peak height and width values at each m/z ratio as a function of time are stored in a memory. A mass spectrometer operation on a material to be analyzed is performed and the peak height and width values at each m/z ratio versus time are stored in a second memory location. The mass spectrometer operation on the material to be analyzed is repeated a fixed number of times and the stored control sample values at each m/z ratio level at each time increment are subtracted from each corresponding one from the operational runs, thus producing a difference value at each mass ratio for each of the multiple runs at each time increment. If the MS value minus the background noise does not exceed a preset value, the m/z ratio data point is not recorded, thus eliminating background noise, chemical noise and false positive peaks from the mass spectrometer data. The stored data for each of the multiple runs is then compared to a predetermined value at each m/z ratio and the resultant series of peaks, which are now determined to be above the background, is stored in the m/z points in which the peaks are of significance.

One possibility for the utilization of mass spectrometry in high throughput screening is the integration of SPR with mass spectrometry. Approaches that have been tried are direct analysis of the analyte retained on the sensor chip and mass spectrometry with the eluted analyte (Sonksen *et al.*, 1998, Anal. Chem. 70:2731-2736; Nelson & Krone, 1999, J. Mol. Recog. 12:77-93). Further developments, especially in the interfacing of the sensor chip with the mass spectrometer and in reusing the sensor chip, are required to make SPR combined with mass spectroscopy a high-throughput method for biomolecular interaction analysis and the screening of targets for small molecule inhibitors (Xavier *et al.*, 2000, Trends Biotechnol. 18(8):349-356).

In one embodiment of the present invention, the target nucleic acid complexed to a test compound can be determined by any of the mass spectrometry processed described *supra*. Furthermore, mass spectrometry can also be used to elucidate

the structure of the test compound.

5.5.5. Scintillation Proximity Assay ("SPA")

5 Scintillation Proximity Assay ("SPA") is a method that can be used for screening small molecules that bind to the target RNAs. SPA would involve radiolabeling either the target RNA or the test compound and then quantitating its binding to the other member to a bead or a surface impregnated with a scintillant (Cook, 1996, Drug Discov. Today 1:287-294). Currently, fluorescence-based techniques are preferred for high-throughput screening (Pope *et al.*, 1999, Drug Discov. Today 4:350-362).

10 Screening for small molecules that inhibit Tat peptide:TAR RNA interaction has been performed with SPA, and inhibitors of the interaction were isolated and characterized (Mei *et al.*, 1997, Bioorg. Med. Chem. 5:1173-1184; Mei *et al.*, 1998, Biochemistry 37:14204-14212). A similar approach can be used to identify small molecules that directly bind to a preselected target RNA element in accordance with the invention can be utilized.

SPA can be adapted to high throughput screening by the availability of microplates, wherein the scintillant is directly incorporated into the plastic of the microtiter wells (Nakayama *et al.*, 1998, J. Biomol. Screening 3:43-48). Thus, one embodiment of the present invention comprises (a) labeling of the target nucleic acid with a radioactive or fluorescent label; (b) contacted the labeled nucleic acid with test compounds, wherein each test compound is in a microtiter well coated with scintillant and is tethered to the microtiter well; and (c) identifying and quantifying the test compounds bound to the target nucleic acid with SPA, wherein the test compound is identified by virtue of its location in the microplate.

5.5.6. Structure-Activity Relationships ("SAR") by NMR Spectroscopy

NMR spectroscopy is a valuable technique for identifying complexed target nucleic acids by qualitatively determining changes in chemical shift, specifically from distances measured using relaxation effects, and NMR-based approaches have been used in the identification of small molecule binders of protein drug targets (Xavier *et al.*, 2000, Trends Biotechnol. 18(8):349-356). The determination of structure-activity relationships ("SAR") by NMR is the first method for NMR described in which small molecules that bind adjacent subsites are identified by two-dimensional ^1H - ^{15}N spectra of the target protein (Shuker *et al.*, 1996, Science 274:1531-1534). The signal from the bound molecule is monitored by employing line broadening, transferred NOEs and pulsed field gradient

diffusion measurements (Moore, 1999, Curr. Opin. Biotechnol. 10:54-58). A strategy for lead generation by NMR using a library of small molecules has been recently described (Fejzo *et al.*, 1999, Chem. Biol. 6:755-769).

- 5 In one embodiment of the present invention, the target nucleic acid complexed to a test compound can be determined by SAR by NMR. Furthermore, SAR by NMR can also be used to elucidate the structure of the test compound.

5.5.7. Size Exclusion Chromatography

- 10 In another embodiment of the present invention, size-exclusion chromatography is used to purify test compounds that are bound to a target nucleic from a complex mixture of compounds. Size-exclusion chromatography separates molecules based on their size and uses gel-based media comprised of beads with specific size distributions. When applied to a column, this media settles into a tightly packed matrix and
15 forms a complex array of pores. Separation is accomplished by the inclusion or exclusion of molecules by these pores based on molecular size. Small molecules are included into the pores and, consequently, their migration through the matrix is retarded due to the added distance they must travel before elution. Large molecules are excluded from the pores and migrate with the void volume when applied to the matrix. In the present invention, a target
20 nucleic acid is incubated with a mixture of test compounds while free in solution and allowed to reach equilibrium. When applied to a size exclusion column, test compounds free in solution are retained by the column, and test compounds bound to the target nucleic acid are passed through the column. In a preferred embodiment, spin columns commonly used for "desalting" of nucleic acids will be employed to separate bound from unbound test
25 compounds (*e.g.*, Bio-Spin columns manufactured by BIO-RAD). In another embodiment, the size exclusion matrix is packed into multiwell plates to allow high throughput separation of mixtures (*e.g.*, PLASMID 96-well SEC plates manufactured by Millipore).

5.5.8. Affinity Chromatography

- 30 In one embodiment of the present invention, affinity capture is used to purify test compounds that are bound to a target nucleic acid labeled with an affinity tag from a complex mixture of compounds. To accomplish this, a target nucleic acid labeled with an affinity tag is incubated with a mixture of test compounds while free in solution and then captured to a solid support once equilibrium has been established; alternatively, target
35 nucleic acids labeled with an affinity tag can be captured to a solid support first and then allowed to reach equilibrium with a mixture of test compounds.

The solid support is typically comprised of, but not limited to, cross-linked agarose beads that are coupled with a ligand for the affinity tag. Alternatively, the solid support may be a glass, silicon, metal, or carbon, plastic (polystyrene, polypropylene) surface with or without a self-assembled monolayer (SAM) either with a covalently attached ligand for the affinity tag, or with inherent affinity for the tag on the target nucleic acid.

Once the complex between the target nucleic acid and test compound has reached equilibrium and has been captured, one skilled in the art will appreciate that the retention of bound compounds and removal of unbound compounds is facilitated by washing the solid support with large excesses of binding reaction buffer. Furthermore, retention of high affinity compounds and removal of low affinity compounds can be accomplished by a number of means that increase the stringency of washing; these means include, but are not limited to, increasing the number and duration of washes, raising the salt concentration of the wash buffer, addition of detergent or surfactant to the wash buffer, and addition of non-specific competitor to the wash buffer.

In one embodiment, the test compounds themselves are detectably labeled with fluorescent dyes, radioactive isotopes, or nanoparticles. When the test compounds are applied to the captured target nucleic acid in a spatially addressed fashion (*e.g.*, in separate wells of a 96-well microplate), binding between the test compounds and the target nucleic acid can be determined by the presence of the detectable label on the test compound using fluorescence.

Following the removal of unbound compounds, bound compounds with high affinity for the target nucleic acid can be eluted from the immobilized target nucleic acids and analyzed. The elution of test compounds can be accomplished by any means that break the non-covalent interactions between the target nucleic acid and compound. Means for elution include, but are not limited to, changing the pH, changing the salt concentration, the application of organic solvents, and the application of molecules that compete with the bound ligand. In a preferred embodiment, the means employed for elution will release the compound from the target RNA, but will not effect the interaction between the affinity tag and the solid support, thereby achieving selective elution of test compound. Moreover, a preferred embodiment will employ an elution buffer that is volatile to allow for subsequent concentration by lyophilization of the eluted compound (*e.g.*, 0 M to 5 M ammonium acetate).

5.5.9. Nanoparticle Aggregation

In one embodiment of the present invention, both the target nucleic acid and the test compounds are labeled with nanoparticles. A nanoparticle is a cluster of ions with controlled size from 0.1 to 1000 nm comprised of metals, metal oxides, or semiconductors including, but not limited to Ag_2S , ZnS , CdS , CdTe , Au , or TiO_2 . Methods for the attachment of nucleic acids and small molecules to nanoparticles are well known to one of skill in the art (reviewed in Niemeyer, 2001, *Angew. Chem. Int. Ed.* 40:4129-4158. The references cited therein are hereby incorporated by reference in their entireties). In particular, if multiple copies of the target nucleic acid are attached to a single nanoparticle and multiple copies of a test compound are attached to another nanoparticle, then interaction between the test compound and target nucleic acid will induce aggregation of nanoparticles as described, for example, by Mitchel *et al.* 1999, *J. Am. Chem. Soc.* 121:8122-8123. The aggregate can be detected by changes in absorbance or fluorescence spectra and physically separated from the unbound components through filtration or centrifugation.

5.6. Methods for Identifying or Characterizing the Test Compounds Bound to the Target Nucleic Acids

If the library comprises arrays or microarrays of test compounds, wherein each test compound has an address or identifier, the test compound can be deconvoluted, *e.g.*, by cross-referencing the positive sample to original compound list that was applied to the individual test assays.

If the library is a peptide or nucleic acid library, the sequence of the test compound can be determined by direct sequencing of the peptide or nucleic acid. Such methods are well known to one of skill in the art.

A number of physico-chemical techniques can be used for the de novo characterization of test compounds bound to the target.

5.6.1. Mass Spectrometry

Mass spectrometry (*e.g.*, electrospray ionization ("ESI") and matrix-assisted laser desorption-ionization ("MALDI"), Fourier-transform ion cyclotron resonance ("FT-ICR")) can be used both for high-throughput screening of test compounds that bind to a target RNA and elucidating the structure of the test compound. Thus, one example of mass spectroscopy is that separation of a bound and unbound complex and test compound structure elucidation can be carried out in a single step.

MALDI uses a pulsed laser for desorption of the ions and a time-of-flight analyzer, and has been used for the detection of noncovalent tRNA:amino-acyl-tRNA synthetase complexes (Gruic-Sovolj *et al.*, 1997, J. Biol. Chem. 272:32084-32091).

5 However, covalent cross-linking between the target nucleic acid and the test compound is required for detection, since a non-covalently bound complex may dissociate during the MALDI process.

ESI mass spectrometry ("ESI-MS") has been of greater utility for studying non-covalent molecular interactions because, unlike the MALDI process, ESI-MS generates molecular ions with little to no fragmentation (Xavier *et al.*, 2000, Trends Biotechnol. 18(8):349-356). ESI-MS has been used to study the complexes formed by HIV Tat peptide and protein with the TAR RNA (Sannes-Lowery *et al.*, 1997, Anal. Chem. 69:5130-5135).

Fourier-transform ion cyclotron resonance ("FT-ICR") mass spectrometry provides high-resolution spectra, isotope-resolved precursor ion selection, and accurate mass assignments (Xavier *et al.*, 2000, Trends Biotechnol. 18(8):349-356). FT-ICR has been used to study the interaction of aminoglycoside antibiotics with cognate and non-cognate RNAs (Hofstadler *et al.*, 1999, Anal. Chem. 71:3436-3440; Griffey *et al.*, 1999, Proc. Natl. Acad. Sci. USA 96:10129-10133). As true for all of the mass spectrometry methods discussed herein, FT-ICR does not require labeling of the target RNA or a test compound.

20 An advantage of mass spectroscopy is not only the elucidation of the structure of the test compound, but also the determination of the structure of the test compound bound to the preselected target RNA. Such information can enable the discovery of a consensus structure of a test compound that specifically binds to a preselected target RNA.

5.6.2. NMR Spectroscopy

As described above, NMR spectroscopy is a technique for identifying binding sites in target nucleic acids by qualitatively determining changes in chemical shift, specifically from distances measured using relaxation effects. Examples of NMR that can be used for the invention include, but are not limited to, one-dimensional NMR, two-dimensional NMR, correlation spectroscopy ("COSY"), and nuclear Overhauser effect ("NOE") spectroscopy. Such methods of structure determination of test compounds are well known to one of skill in the art.

35 Similar to mass spectroscopy, an advantage of NMR is the not only the elucidation of the structure of the test compound, but also the determination of the structure

of the test compound bound to the preselected target RNA. Such information can enable the discovery of a consensus structure of a test compound that specifically binds to a preselected target RNA.

5

5.6.3. Vibrational Spectroscopy

Vibrational spectroscopy (*e.g.* infrared (IR) spectroscopy or Raman spectroscopy) can be used for elucidating the structure of the test compound on the isolated bead.

10

Infrared spectroscopy measures the frequencies of infrared light (wavelengths from 100 to 10,000 nm) absorbed by the test compound as a result of excitation of vibrational modes according to quantum mechanical selection rules which require that absorption of light cause a change in the electric dipole moment of the molecule. The infrared spectrum of any molecule is a unique pattern of absorption wavelengths of varying intensity that can be considered as a molecular fingerprint to identify any compound.

15

Infrared spectra can be measured in a scanning mode by measuring the absorption of individual frequencies of light, produced by a grating which separates frequencies from a mixed-frequency infrared light source, by the test compound relative to a standard intensity (double-beam instrument) or pre-measured ('blank') intensity (single-beam instrument). In a preferred embodiment, infrared spectra are measured in a pulsed mode (FT-IR) where a mixed beam, produced by an interferometer, of all infrared light frequencies is passed through or reflected off the test compound. The resulting interferogram, which may or may not be added with the resulting interferograms from subsequent pulses to increase the signal strength while averaging random noise in the electronic signal, is mathematically transformed into a spectrum using Fourier Transform or Fast Fourier Transform algorithms.

20

25

Raman spectroscopy measures the difference in frequency due to absorption of infrared frequencies of scattered visible or ultraviolet light relative to the incident beam.

30

The incident monochromatic light beam, usually a single laser frequency, is not truly absorbed by the test compound but interacts with the electric field transiently. Most of the light scattered off the sample will be unchanged (Rayleigh scattering) but a portion of the scatter light will have frequencies that are the sum or difference of the incident and molecular vibrational frequencies. The selection rules for Raman (inelastic) scattering require a change in polarizability of the molecule. While some vibrational transitions are observable in both infrared and Raman spectrometry, must be observable only with one or

35

the other technique. The Raman spectrum of any molecule is a unique pattern of absorption wavelengths of varying intensity that can be considered as a molecular fingerprint to identify any compound.

5 Raman spectra are measured by submitting monochromatic light to the sample, either passed through or preferably reflected off, filtering the Rayleigh scattered light, and detecting the frequency of the Raman scattered light. An improved Raman spectrometer is described in US Patent No. 5,786,893 to Fink *et al.*, which is hereby incorporated by reference.

10 Vibrational microscopy can be measured in a spatially resolved fashion to address single beads by integration of a visible microscope and spectrometer. A microscopic infrared spectrometer is described in U.S. Patent No. 5,581,085 to Refrner *et al.*, which is hereby incorporated by reference in its entirety. An instrument that simultaneously performs a microscopic infrared and microscopic Raman analysis on a sample is described in U.S. Patent No. 5,841,139 to Sostek *et al.*, which is hereby
15 incorporated by reference in its entirety.

In the preferred embodiment, test compounds can be identified by matching the IR or Raman spectra of a test compound to a dataset of vibrational (IR or Raman) spectra previously acquired for each compound in the combinatorial library. By this
20 method, the spectra of compounds with known structure are recorded so that comparison with these spectra can identify compounds again when isolated from RNA binding experiments.

5.7. Secondary Biological Screens

25 The test compounds identified in the binding assay (for convenience referred to herein as a "lead" compound) can be tested for biological activity using host cells containing or engineered to contain the target RNA element coupled to a functional readout system. For example, the lead compound can be tested in a host cell engineered to contain the target RNA element controlling the expression of a reporter gene. In this example, the
30 lead compounds are assayed in the presence or absence of the target RNA. Alternatively, a phenotypic or physiological readout can be used to assess activity of the target RNA in the presence and absence of the lead compound.

In one embodiment, the lead compound can be tested in a host cell engineered to contain the target RNA element controlling the expression of a reporter gene,
35 such as, but not limited to, β -galactosidase, green fluorescent protein, red fluorescent protein, luciferase, chloramphenicol acetyltransferase, alkaline phosphatase, and β -

lactamase. In a preferred embodiment, a cDNA encoding the target element is fused upstream to a reporter gene wherein translation of the reporter gene is repressed upon binding of the lead compound to the target RNA. In other words, the steric hindrance caused by the binding of the lead compound to the target RNA repressed the translation of the reporter gene. This method, termed the translational repression assay procedure ("TRAP") has been demonstrated in *E. coli* and *S. cerevisiae* (Jain & Belasco, 1996, Cell 87(1):115-25; Huang & Schreiber, 1997, Proc. Natl. Acad. Sci. USA 94:13396-13401).

In another embodiment, a phenotypic or physiological readout can be used to assess activity of the target RNA in the presence and absence of the lead compound. For example, the target RNA may be overexpressed in a cell in which the target RNA is endogenously expressed. Where the target RNA controls expression of a gene product involved in cell growth or viability, the *in vivo* effect of the lead compound can be assayed by measuring the cell growth or viability of the target cell. Alternatively, a reporter gene can also be fused downstream of the target RNA sequence and the effect of the lead compound on reporter gene expression can be assayed.

Alternatively, the lead compounds identified in the binding assay can be tested for biological activity using animal models for a disease, condition, or syndrome of interest. These include animals engineered to contain the target RNA element coupled to a functional readout system, such as a transgenic mouse. Animal model systems can also be used to demonstrate safety and efficacy.

Compounds displaying the desired biological activity can be considered to be lead compounds, and will be used in the design of congeners or analogs possessing useful pharmacological activity and physiological profiles. Following the identification of a lead compound, molecular modeling techniques can be employed, which have proven to be useful in conjunction with synthetic efforts, to design variants of the lead that can be more effective. These applications may include, but are not limited to, Pharmacophore Modeling (*cf.* Lamothe, *et al.* 1997, J. Med. Chem. 40: 3542; Mottola *et al.* 1996, J. Med. Chem. 39: 285; Beusen *et al.* 1995, Biopolymers 36: 181; P. Fossa *et al.* 1998, Comput. Aided Mol. Des. 12: 361), QSAR development (*cf.* Siddiqui *et al.* 1999, J. Med. Chem. 42: 4122; Barreca *et al.* 1999 Bioorg. Med. Chem. 7: 2283; Kroemer *et al.* 1995, J. Med. Chem. 38: 4917; Schaal *et al.* 2001, J. Med. Chem. 44: 155; Buolamwini & Assefa 2002, J. Mol. Chem. 45: 84), Virtual docking and screening/scoring (*cf.* Anzini *et al.* 2001, J. Med. Chem. 44: 1134; Faaland *et al.* 2000, Biochem. Cell. Biol. 78: 415; Silvestri *et al.* 2000, Bioorg. Med. Chem. 8: 2305; J. Lee *et al.* 2001, Bioorg. Med. Chem. 9: 19), and Structure Prediction using RNA structural programs including, but not limited to mFold (as described

by Zuker *et al.* Algorithms and Thermodynamics for RNA Secondary Structure Prediction: A Practical Guide in RNA Biochemistry and Biotechnology pp. 11-43, J. Barciszewski & B.F.C. Clark, eds. (NATO ASI Series, Kluwer Academic Publishers, 1999) and Mathews *et al.* 1999 J. Mol. Biol. 288: 911-940); RNAmotif (Macke *et al.* 2001, Nucleic Acids Res. 29: 4724-4735; and the Vienna RNA package (Hofacker *et al.* 1994, Monatsh. Chem. 125: 167-188).

Further examples of the application of such techniques can be found in several review articles, such as Rotivinen *et al.*, 1988, Acta Pharmaceutica Fennica 97:159-166; Ripka, 1998, New Scientist 54-57; McKinaly & Rossmann, 1989, Annu. Rev. Pharmacol. Toxicol. 29:111-122; Perry & Davies, QSAR: Quantitative Structure-Activity Relationships in Drug Design pp. 189-193 (Alan R. Liss, Inc. 1989); Lewis & Dean, 1989, Proc. R. Soc. Lond. 236:125-140 and 141-162; Askew *et al.*, 1989, J. Am. Chem. Soc. 111:1082-1090. Molecular modeling tools employed may include those from Tripos, Inc., St. Louis, Missouri (e.g., Sybyl/UNITY, CONCORD, DiverseSolutions), Accelrys, San Diego, California (e.g., Catalyst, Wisconsin Package (BLAST, etc.)), Schrodinger, Portland, Oregon (e.g., QikProp, QikFit, Jaguar) or other such vendors as BioDesign, Inc. (Pasadena, California), Allelix, Inc. (Mississauga, Ontario, Canada), and Hypercube, Inc. (Cambridge, Ontario, Canada), and may include privately designed and/or "academic" software (e.g. RNAmotif, mFOLD). These application suites and programs include tools for the atomistic construction and analysis of structural models for drug-like molecules, proteins, and DNA or RNA and their potential interactions. They also provide for the calculation of important physical properties, such as solubility estimates, permeability metrics, and empirical measures of molecular "druggability" (e.g., Lipinski "Rule of 5" as described by Lipinski *et al.* 1997, Adv. Drug Delivery Rev. 23: 3-25). Most importantly, they provide appropriate metrics and statistical modeling power (such as the patented CoMFA technology in Sybyl as described in US Patents 6,240,374 and 6,185,506) to develop Quantitative Structural Activity Relationships (QSARs) which are used to guide the synthesis of more efficacious clinical development candidates while improving desirable physical properties, as determined by results from the aforementioned secondary screening protocols.

5.8. Use of Identified Compounds That Bind RNA to Treat/Prevent Disease

Biologically active compounds identified using the methods of the invention or a pharmaceutically acceptable salt thereof can be administered to a patient, preferably a mammal, more preferably a human, suffering from a disease whose progression is

associated with a target RNA:host cell factor interaction *in vivo*. In certain embodiments, such compounds or a pharmaceutically acceptable salt thereof is administered to a patient, preferably a mammal, more preferably a human, as a preventative measure against a disease associated with an RNA:host cell factor interaction *in vivo*.

In one embodiment, "treatment" or "treating" refers to an amelioration of a disease, or at least one discernible symptom thereof. In another embodiment, "treatment" or "treating" refers to an amelioration of at least one measurable physical parameter, not necessarily discernible by the patient. In yet another embodiment, "treatment" or "treating" refers to inhibiting the progression of a disease, either physically, *e.g.*, stabilization of a discernible symptom, physiologically, *e.g.*, stabilization of a physical parameter, or both. In yet another embodiment, "treatment" or "treating" refers to delaying the onset of a disease.

In certain embodiments, the compound or a pharmaceutically acceptable salt thereof is administered to a patient, preferably a mammal, more preferably a human, as a preventative measure against a disease associated with an RNA:host cell factor interaction *in vivo*. As used herein, "prevention" or "preventing" refers to a reduction of the risk of acquiring a disease. In one embodiment, the compound or a pharmaceutically acceptable salt thereof is administered as a preventative measure to a patient. According to this embodiment, the patient can have a genetic predisposition to a disease, such as a family history of the disease, or a non-genetic predisposition to the disease. Accordingly, the compound and pharmaceutically acceptable salts thereof can be used for the treatment of one manifestation of a disease and prevention of another.

When administered to a patient, the compound or a pharmaceutically acceptable salt thereof is preferably administered as component of a composition that optionally comprises a pharmaceutically acceptable vehicle. The composition can be administered orally, or by any other convenient route, for example, by infusion or bolus injection, by absorption through epithelial or mucocutaneous linings (*e.g.*, oral mucosa, rectal, and intestinal mucosa, *etc.*) and may be administered together with another biologically active agent. Administration can be systemic or local. Various delivery systems are known, *e.g.*, encapsulation in liposomes, microparticles, microcapsules, capsules, *etc.*, and can be used to administer the compound and pharmaceutically acceptable salts thereof.

Methods of administration include but are not limited to intradermal, intramuscular, intraperitoneal, intravenous, subcutaneous, intranasal, epidural, oral, sublingual, intranasal, intracerebral, intravaginal, transdermal, rectally, by inhalation, or topically, particularly to the ears, nose, eyes, or skin. The mode of administration is left to

the discretion of the practitioner. In most instances, administration will result in the release of the compound or a pharmaceutically acceptable salt thereof into the bloodstream.

In specific embodiments, it may be desirable to administer the compound or a pharmaceutically acceptable salt thereof locally. This may be achieved, for example, and not by way of limitation, by local infusion during surgery, topical application, *e.g.*, in conjunction with a wound dressing after surgery, by injection, by means of a catheter, by means of a suppository, or by means of an implant, said implant being of a porous, non-porous, or gelatinous material, including membranes, such as sialastic membranes, or fibers.

In certain embodiments, it may be desirable to introduce the compound or a pharmaceutically acceptable salt thereof into the central nervous system by any suitable route, including intraventricular, intrathecal and epidural injection. Intraventricular injection may be facilitated by an intraventricular catheter, for example, attached to a reservoir, such as an Ommaya reservoir.

Pulmonary administration can also be employed, *e.g.*, by use of an inhaler or nebulizer, and formulation with an aerosolizing agent, or via perfusion in a fluorocarbon or synthetic pulmonary surfactant. In certain embodiments, the compound and pharmaceutically acceptable salts thereof can be formulated as a suppository, with traditional binders and vehicles such as triglycerides.

In another embodiment, the compound and pharmaceutically acceptable salts thereof can be delivered in a vesicle, in particular a liposome (see Langer, 1990, *Science* 249:1527-1533; Treat *et al.*, in *Liposomes in the Therapy of Infectious Disease and Cancer*, Lopez-Berestein and Fidler (eds.), Liss, New York, pp. 353-365 (1989); Lopez-Berestein, *ibid.*, pp. 317-327; see generally *ibid.*).

In yet another embodiment, the compound and pharmaceutically acceptable salts thereof can be delivered in a controlled release system (see, *e.g.*, Goodson, in *Medical Applications of Controlled Release*, supra, vol. 2, pp. 115-138 (1984)). Other controlled-release systems discussed in the review by Langer, 1990, *Science* 249:1527-1533) may be used. In one embodiment, a pump may be used (see Langer, supra; Sefton, 1987, *CRC Crit. Ref. Biomed. Eng.* 14:201; Buchwald *et al.*, 1980, *Surgery* 88:507 Saudek *et al.*, 1989, *N. Engl. J. Med.* 321:574). In another embodiment, polymeric materials can be used (see *Medical Applications of Controlled Release*, Langer and Wise (eds.), CRC Pres., Boca Raton, Florida (1974); *Controlled Drug Bioavailability, Drug Product Design and Performance*, Smolen and Ball (eds.), Wiley, New York (1984); Ranger and Peppas, 1983, *J. Macromol. Sci. Rev. Macromol. Chem.* 23:61; see also Levy *et al.*, 1985, *Science*

228:190; During *et al.*, 1989, *Ann. Neurol.* 25:351; Howard *et al.*, 1989, *J. Neurosurg.* 71:105). In yet another embodiment, a controlled-release system can be placed in proximity of a target RNA of the compound or a pharmaceutically acceptable salt thereof, thus requiring only a fraction of the systemic dose.

Compositions comprising the compound or a pharmaceutically acceptable salt thereof ("compound compositions") can additionally comprise a suitable amount of a pharmaceutically acceptable vehicle so as to provide the form for proper administration to the patient.

In a specific embodiment, the term "pharmaceutically acceptable" means approved by a regulatory agency of the Federal or a state government or listed in the U.S. Pharmacopeia or other generally recognized pharmacopeia for use in animals, mammals, and more particularly in humans. The term "vehicle" refers to a diluent, adjuvant, excipient, or carrier with which a compound of the invention is administered. Such pharmaceutical vehicles can be liquids, such as water and oils, including those of petroleum, animal, vegetable or synthetic origin, such as peanut oil, soybean oil, mineral oil, sesame oil and the like. The pharmaceutical vehicles can be saline, gum acacia, gelatin, starch paste, talc, keratin, colloidal silica, urea, and the like. In addition, auxiliary, stabilizing, thickening, lubricating and coloring agents may be used. When administered to a patient, the pharmaceutically acceptable vehicles are preferably sterile. Water is a preferred vehicle when the compound of the invention is administered intravenously. Saline solutions and aqueous dextrose and glycerol solutions can also be employed as liquid vehicles, particularly for injectable solutions. Suitable pharmaceutical vehicles also include excipients such as starch, glucose, lactose, sucrose, gelatin, malt, rice, flour, chalk, silica gel, sodium stearate, glycerol monostearate, talc, sodium chloride, dried skim milk, glycerol, propylene, glycol, water, ethanol and the like. Compound compositions, if desired, can also contain minor amounts of wetting or emulsifying agents, or pH buffering agents.

Compound compositions can take the form of solutions, suspensions, emulsion, tablets, pills, pellets, capsules, capsules containing liquids, powders, sustained-release formulations, suppositories, emulsions, aerosols, sprays, suspensions, or any other form suitable for use. In one embodiment, the pharmaceutically acceptable vehicle is a capsule (see *e.g.*, U.S. Patent No. 5,698,155). Other examples of suitable pharmaceutical vehicles are described in Remington's Pharmaceutical Sciences, Alfonso R. Genmaro, ed., Mack Publishing Co. Easton, PA, 19th ed., 1995, pp. 1447 to 1676, incorporated herein by reference.

In a preferred embodiment, the compound or a pharmaceutically acceptable salt thereof is formulated in accordance with routine procedures as a pharmaceutical composition adapted for oral administration to human beings. Compositions for oral delivery may be in the form of tablets, lozenges, aqueous or oily suspensions, granules, powders, emulsions, capsules, syrups, or elixirs, for example. Orally administered compositions may contain one or more agents, for example, sweetening agents such as fructose, aspartame or saccharin; flavoring agents such as peppermint, oil of wintergreen, or cherry; coloring agents; and preserving agents, to provide a pharmaceutically palatable preparation. Moreover, where in tablet or pill form, the compositions can be coated to delay disintegration and absorption in the gastrointestinal tract thereby providing a sustained action over an extended period of time. Selectively permeable membranes surrounding an osmotically active driving compound are also suitable for orally administered compositions. In these later platforms, fluid from the environment surrounding the capsule is imbibed by the driving compound, which swells to displace the agent or agent composition through an aperture. These delivery platforms can provide an essentially zero order delivery profile as opposed to the spiked profiles of immediate release formulations. A time delay material such as glycerol monostearate or glycerol stearate may also be used. Oral compositions can include standard vehicles such as mannitol, lactose, starch, magnesium stearate, sodium saccharine, cellulose, magnesium carbonate, and the like. Such vehicles are preferably of pharmaceutical grade. Typically, compositions for intravenous administration comprise sterile isotonic aqueous buffer. Where necessary, the compositions may also include a solubilizing agent.

In another embodiment, the compound or a pharmaceutically acceptable salt thereof can be formulated for intravenous administration. Compositions for intravenous administration may optionally include a local anesthetic such as lignocaine to lessen pain at the site of the injection. Generally, the ingredients are supplied either separately or mixed together in unit dosage form, for example, as a dry lyophilized powder or water-free concentrate in a hermetically sealed container such as an ampoule or sachette indicating the quantity of active agent. Where the compound or a pharmaceutically acceptable salt thereof is to be administered by infusion, it can be dispensed, for example, with an infusion bottle containing sterile pharmaceutical grade water or saline. Where the compound or a pharmaceutically acceptable salt thereof is administered by injection, an ampoule of sterile water for injection or saline can be provided so that the ingredients may be mixed prior to administration.

The amount of a compound or a pharmaceutically acceptable salt thereof that will be effective in the treatment of a particular disease will depend on the nature of the disease, and can be determined by standard clinical techniques. In addition, *in vitro* or *in vivo* assays may optionally be employed to help identify optimal dosage ranges. The precise dose to be employed will also depend on the route of administration, and the seriousness of the disease, and should be decided according to the judgment of the practitioner and each patient's circumstances. However, suitable dosage ranges for oral administration are generally about 0.001 milligram to about 200 milligrams of a compound or a pharmaceutically acceptable salt thereof per kilogram body weight per day. In specific preferred embodiments of the invention, the oral dose is about 0.01 milligram to about 100 milligrams per kilogram body weight per day, more preferably about 0.1 milligram to about 75 milligrams per kilogram body weight per day, more preferably about 0.5 milligram to 5 milligrams per kilogram body weight per day. The dosage amounts described herein refer to total amounts administered; that is, if more than one compound is administered, or if a compound is administered with a therapeutic agent, then the preferred dosages correspond to the total amount administered. Oral compositions preferably contain about 10% to about 95% active ingredient by weight.

Suitable dosage ranges for intravenous (i.v.) administration are about 0.01 milligram to about 100 milligrams per kilogram body weight per day, about 0.1 milligram to about 35 milligrams per kilogram body weight per day, and about 1 milligram to about 10 milligrams per kilogram body weight per day. Suitable dosage ranges for intranasal administration are generally about 0.01 pg/kg body weight per day to about 1 mg/kg body weight per day. Suppositories generally contain about 0.01 milligram to about 50 milligrams of a compound of the invention per kilogram body weight per day and comprise active ingredient in the range of about 0.5% to about 10% by weight.

Recommended dosages for intradermal, intramuscular, intraperitoneal, subcutaneous, epidural, sublingual, intracerebral, intravaginal, transdermal administration or administration by inhalation are in the range of about 0.001 milligram to about 200 milligrams per kilogram of body weight per day. Suitable doses for topical administration are in the range of about 0.001 milligram to about 1 milligram, depending on the area of administration. Effective doses may be extrapolated from dose-response curves derived from *in vitro* or animal model test systems. Such animal models and systems are well known in the art.

The compound and pharmaceutically acceptable salts thereof are preferably assayed *in vitro* and *in vivo*, for the desired therapeutic or prophylactic activity, prior to use

in humans. For example, *in vitro* assays can be used to determine whether it is preferable to administer the compound, a pharmaceutically acceptable salt thereof, and/or another therapeutic agent. Animal model systems can be used to demonstrate safety and efficacy.

5 A variety of compounds can be used for treating or preventing diseases in mammals. Types of compounds include, but are not limited to, peptides, peptide analogs including peptides comprising non-natural amino acids, *e.g.*, D-amino acids, phosphorous analogs of amino acids, such as α -amino phosphonic acids and α -amino phosphinic acids, or amino acids having non-peptide linkages, nucleic acids, nucleic acid analogs such as
10 phosphorothioates or peptide nucleic acids ("PNAs"), hormones, antigens, synthetic or naturally occurring drugs, opiates, dopamine, serotonin, catecholamines, thrombin, acetylcholine, prostaglandins, organic molecules, pheromones, adenosine, sucrose, glucose, lactose and galactose.

15 6. EXAMPLE: THERAPEUTIC TARGETS

The therapeutic targets presented herein are by way of example, and the present invention is not to be limited by the targets described herein. The therapeutic targets presented herein as DNA sequences are understood by one of skill in the art that the sequences can be converted to RNA sequences.

20 6.1. Tumor Necrosis Factor Alpha ("TNF- α ")

GenBank Accession # X01394:

1 gcagaggacc agctaagagg gagagaagca actacagacc cccctgaaa acaacctca
61 gacgccacat cccctgacaa gctgccaggc aggttctctt cctctcacat actgaccac
25 121 ggctccaccc tctctccct ggaaggaca ccatgagcac tgaagcatg atccgggagc
181 tggagctggc cgaggaggcg ctcccacaaga agacaggggg gccccaggcg tccaggcggt
241 gctgttctct cagcctcttc tcttctctga tctgtgcagg cgcaccacg cttctctgc
301 tctgctactt tggagtgtac ggcgccccaga ggaagagtt cccaggcgac ctctctctaa
361 tcagccctct ggcccaggca gtcgatcat cttctgaac ccgagtgac aagcctgtag
421 cccatgttgt agcaaacctt caagctgagg ggcagctcca gtggctgaac cgccggggca
481 atgccctctt ggccaatggc gtggagctga gagaatacca gctgggtgtg ccatcagagg
541 gccgtacct catctactcc caggtcctct tcaaggccca aggcgtgcc cccaccatg
601 tctctctcac ccacaccatc agccgcacgc cgtctccta ccagaccaag gtcaacctcc
661 tctctgcat caagagcccc tggcagaggg agaccgccga gggggctgag gccaaagccct
721 ggtatgagcc catctatctg ggaagggtct tccagctgga gaagggtgac cgactcagcg
35 781 ctgagatcaa tggcccgcac tatctcgact ttgccgagtc tgggcaggtc tactttggga

841 tcattgccct gtgaggagga cgaacatcca accttcccaa acgcctcccc tgcccacaac
 901 ccttattac cccctcttc agacaccctc aacctctct ggctcaaaaa gagaattggg
 961 ggcttagggc cggaacccaa gcttagaact ttaagcaaca agaccaccac ttcgaaact
 1021 gggattcagg aatgtgtgct ctgcacagtg aattgctggc aaccactaag aattcaaac
 1081 ggggctcca gaactactg gggcctacag ctttgatccc tgacatctgg aatctggaga
 1141 ccaggggagc ttgtgtctg gccagaatgc tgcaggactt gagaagacct cacctagaaa
 1201 ttgacacaag tggaccttag gccctctct ctcagatgt ttcagactt ccttgagaca
 1261 cggagcccag cctccccc ggagccagct cctctattt atgtttgac ttgtgattat
 1321 ttattatta ttattattt attatttac agatgaatgt aittatttg gagaccgggg
 1381 tatctgggg gacccaatgt aggagctgcc ttggctcaga catgttttc gtgaaaacgg
 1441 agctgaacaa taggtgttc ccatgtagcc ccttggctc tgtgctctt ttgattatg
 1501 tttttaaaa tattattctg attaatgtt ctaaacaaatg ctgatttgg gaccaactgt
 1561 cactcatgce tgagcctctg ctcccagggt gagtgtgtc tgaatgcc ctactattca
 1621 gtggcgagaa ataaagtgg ctt (SEQ ID NO: 6)

General Target Regions:

- (1) 5' Untranslated Region - nts 1 - 152
- (2) 3' Untranslated Region - nts 852 - 1643

Initial Specific Target Motif:

Group I AU-Rich Element (ARE) Cluster in 3' untranslated region
 5' AUUUAUUUAUUUAUUUAUUUA 3' (SEQ ID NO: 1)

6.2. Granulocyte-macrophage Colony Stimulating Factor ("GM-CSF")

GenBank Accession # NM_000758:

1 gctggaggat gtggctcag agcctgtgc tcttgggcac tgtgacctg agcatctctg
 61 caccgcccgc ctgcgccagc ccagcacgc agccctggga gcatglaat gccatccagg
 121 agggccgggc tctctgaac ctgagttag acactgtcgc tgagatgaat gaacagttag
 181 aagtcatctc agaatgttt gacctcagg agccgacctg cctacagacc cgctggagc
 241 tttacaagca gggcctcggg ggcagcctca ccaagctcaa gggcccttg accatgatgg
 301 ccagccacta cagcagcac tgccctcaa ccccggaac ttctgtgca accagacta
 361 tcaccttga aagttcaaa gagaacctga aggactttct gctgtcatc cctttgact
 421 gctgggagcc agtccaggag tgagaccggc cagatgaggc tggccaagcc ggggagctgc
 481 tctctatga aacaagagct agaaactcag gatgtcatc ttggaggac caagggtgg
 541 gccacagcca tgggtggagt ggccctggac tgccctgggc cacactgacc ctgatacagg

601 catggcagaa gaatgggaat attttatact gacagaaatc agtaataatt atatatatt
 661 attttaaaa tatttattta ttttttatt taagtctata ttccataatt atcaagatg
 721 ttttaccgta ataattatta ttaaaaatat gcttct (SEQ ID NO: 7)

5

GenBank Accession # XM_003751:

1 tctggaggat gtggctgcag agcctgctgc tcttgggcac tgtggcctgc agcatctctg
 61 caccgcccg ctcgccagc cccagcacgc agccctggga gcatgtgaat gccatccagg
 121 aggcccgccg tctctgaac ctgagtagag acactgctgc tgagatgaat gaaacagtag
 10 181 aagtcatctc agaatgttt gaactccagg agccgacctg cctacagacc cgcttgagac
 241 tgtacaagca gggcctgcgg ggcagcctca ccaagctcaa gggcccttg accatgatgg
 301 ccagccacta caagcagcac tgcctccaa ccccggaac ttctgtgca accagacta
 361 tcaccttga aagtttcaa gagaactga aggactttt gctgtcatc cctttgact
 421 gctgggagcc agtccaggag tgagaccggc cagatgaggc tggccaagcc ggggagctgc
 15 481 tctctcatga aacaagact agaaactcag gatggcatc ttggagggac caaggggtgg
 541 gccacagcca tgggtggagt ggcctggacc tgcctgggc cacactgacc ctgatacagg
 601 catggcagaa gaatgggaat attttatact gacagaaatc agtaataatt atatatatt
 661 attttaaaa tatttattta ttttttatt taagtctata ttccataatt atcaagatg
 721 ttttaccgta ataattatta ttaaaaatat gcttct (SEQ ID NO: 8)

20

General Target Regions:

- (1) 5' Untranslated Region - nts 1 - 32
- (2) 3' Untranslated Region - nts 468 - 789

25 Initial Specific Target Motif:

Group I AU-Rich Element (ARE) Cluster in 3' untranslated region
 5' AUUUUUUUUUUUUUUUUUUA 3' (SEQ ID NO: 1)

6.3. Interleukin 2 ("IL-2")

30 GenBank Accession # U25676:

1 atcactctct ttaactacta ctacataaa cctcaactc tgcacaaatg tacaggatgc
 61 aactcctgct ttgcatgca ctaattcttg cactgtcac aaacagtga cctacttcaa
 121 gtgcacaaa gaaaacaaag aaacacagc tacaactgga gcatttactg ctggatttac
 181 agatgatttt gaatgggaatt aataattaca agaafcccaa actcaccagg atgtcaccat
 241 ttaagtftta catgcccaag aaggccacag aactgaaca gcttcagtgt ctagaagaag
 35 301 aactcaaacc tctggaggaa gtgctgaatt tagctcaaa caaaaacttt cacttaagac

361 ccagggactt aatcagcaat atcaacgtaa tagttctgga actaaaggga tctgaacaa
 421 caftcatgtg tgaatatgca gatgagacag caaccattgt agaatttctg aacagatgga
 481 ttaccttttg tcaagcatc atcacaac taacttgata attaatgtct tccactflaa
 541 aacalatcag gccttctatt taittattta aatatttaa tttatattt atgttgtaat
 601 gtatggttgc tacctattgt aactattatt cttaacttta aaactataaa tatggatctt
 661 ttatgattct ttttgaagc cctaggggct ctaaaatggt ttacctatt tatcccaaaa
 721 atatttatta ttatgttga ttgtaatat agtatctatg tagatttggt agtaaaaaa
 781 ttaataaat ttgataaata taaaaaaaaa aaacaaaaa aaaaa (SEQ ID NO: 9)

10

General Target Regions:

- (1) 5' Untranslated Region - nts 1 - 47
- (2) 3' Untranslated Region - nts 519- 825

15 Initial Specific Target Motifs:

Group III AU-Rich Element (ARE) Cluster in 3' untranslated region
 5' NAUUUAUUUAUUUAN 3' (SEQ ID NO: 10)

6.4. Interleukin 6 ("IL-6")

20 GenBank Accession # NM_000600:

1 ttctgcctc gagccaccg ggaacgaag agaagctcta tctgcctcc aggagccag
 61 ctatgaatic ctctccaca agcgccttcg gtccagtgc ctctccctg gggtgcctc
 121 tgggtgtgccc tgcctgcctc cctgccccag taccgccagg agaagatcc aaagatgtg
 181 ccgccccaca cagacagcca ctacactctt cagaacgaat tgacaacaaa attcgtgaca
 241 tcttcgacgg catctcagcc ctgagaaagg agacatgtaa caagagtaac atgtgtgaaa
 301 gcagcaaaaga ggcactggca gaaaacaacc tgaacctccc aaagatggct gaaaagaagt
 361 gatcttcca atctggatic aatgaggaga ctgctcgtgt gaaaatcact actggtcttt
 421 tggagtttga ggtataccta ggttacctcc agaacagatt tgagagtagt gaggaacaag
 481 ccagagctgt gcagatgagt acaaaagtc tgatccagtt cctgcagaaa aaggcaaatg
 541 atctagatgc aataaccacc cctgaccocaa ccacaaatgc cagctctgtc acgaagtgc
 601 aggcacagaa ccagtgtgctg caggacatga caactcatct cattctgcgc agctttaagg
 661 agttcctgca gtccagocct agggctcttc ggcaaatgta gcatgggcac ctcagattgt
 721 ttgtgttaat gggcattcct tctctgtgct agaaacctgt ccactgggca cagaacttat
 781 gttgtctct atggagaact aaaagatga gcgttaggac actattttaa ttatttttaa
 841 ttatttaata tttaataatg tgaagctgag ttaatttatg taagtcatat ttattttttt
 901 aagaagtacc acttgaaca ttttatgtat tagttttgaa ataataatgg aaagtggcta

961 tgcagtttga atatcctttg ttccagagcc agatcatttc ttggaaagtg taggcttacc
 1021 tcaataaat ggctaactta tacataatft taaagaaata ttatatattg atttatataa
 1081 tgtataaatg gttttatata caataaatgg cattttaaaa aatic (SEQ ID NO: 11)

5

General Target Regions:

- (1) 5' Untranslated Region - nts 1 - 62
- (2) 3' Untranslated Region - nts 699 - 1125

10

Initial Specific Target Motifs:

Group III AU-Rich Element (ARE) Cluster in 3' untranslated region
 5' NAUUUAUUUAUUUAN 3' (SEQ ID NO: 10)

6.5. Vascular Endothelial Growth Factor ("VEGF")

15 GenBank Accession # AF022375:

1 aagagctcca gagagaagtc gaggaagaga gagacggggt cagagagagc gcgcggggct
 61 gcgagcagcg aaagcgacag gggcaaatg agtgaacctgc ttttgggggt gaccgcggga
 121 gcgcggcggtg agccctcccc ctggggtacc cgcagctgac cagtgcgct gacggacaga
 181 cagacagaca ccgccccag cccagttac cactctctcc ccggccggcg gcggacagtg
 241 gacgcggcgg cgagccggcg gcaggggcgg gagcccgccc ccgagggcgg ggtggagggg
 301 gtccgagctc gcggcgctgc actgaacct ttctccaac ttctgggctg ttctgccttc
 361 ggaggagccg tggctccgcg ggggggaagcc gagccgagcg gagccgcgag aagtgcctagc
 421 tcgggccggg aggagccgca gccggaggag ggggaggagg aagaagagaa ggaagaggag
 481 agggggccgc agtggcgact cggcgctcgg aagccgggt catggacggg tgaggcgcg
 541 gtgtgcgcag acagtgcctc agcgcgcgcg ctccccagcc ctggcccggc ctggggccgg
 601 gaggaagagt agctgcgcca ggccgcgagg agagcgggccc gccccacagc ccgagccgga
 661 gagggacgcg agccgcgcgc cccggctggg cctccgaac calgaactt ctgtctctt
 721 ggtgtcattg gagecttgcc ttgctctct acctccacca tgccaaagtg toccaggctg
 781 caccatggc agaaggagga gggcagatc atcaagaagt ggtgaagttc atggtatgtc
 841 atcagcgagc ctactgccat ccaatcgaga cctgggtgga catctccag gattacctg
 901 atgagatcga gtacatctc aagccatct gtgtgcccct gatcgatgc gggggctgct
 961 ccaatgacga gggcctggag tgtgtgcccc ctgaggagtc caacatcacc atgcagatta
 1021 tgcggatcaa acctaccaa gggccagaca taggagagat gagcttccia cagcacaaca
 1081 aatgtgaatg cagaccaaa gaaagatagag caagacaaga aaatccctgt gggcctgtct
 1141 cagagcggag aaagcatttg ttgtacaag atccgcagag gtgtaaatgt toctgcaaaa
 1201 acacacactc gcgttgcaag gcgaggcagc ttgagttaaa cgaactgact tgcagatgtg

1261 acaagccgag gcggtgagcc ggcgaggagg aaggagcctc cctcagggtt tgggaaacca
 1321 gatctctctc caggaaagac tgatagacaa cgatcgatgc agaaaccacg ctgccggcac
 1381 cacaccatca caatcgacag aacagctctt aatccagaaa cctgaatatg aggaagagga
 5 1441 gactctgcgc agagcacttt gggctcggag ggcgagactc cggcgggaagc attccggggc.
 1501 ggggtgacca gcacggtccc tcttggaatt ggattcgcca tttattttt cttgtcgcta
 1561 aatcacccag cccggaagt tagagagttt taftctggg attctgtag acacaccac
 1621 ccacatacat acatttatat atatatatat tatatatata taaaaataaa tatctctatt
 1681 ttatatatat aaaaatatata taftctttt ttaaattaac agtgctaagt ttattgtgt
 10 1741 cticactgga tgtatttgac tgcgtgac ttgagttggg aggggaatgt tccactcag
 1801 atctcgacag ggaagaggag gagatgagag actctggcat gatcttttt ttgtccact
 1861 tgggtggggc agggctctct cccctgccca agaattgtca agggcagggc atggggggca
 1921 atatgaccca gttttgggaa caccgacaaa cccagccctg gcgctgagcc tctctacccc
 1981 aggtcgacag gacagaaga caaatcacag gtcccgggat gaggcacacc gctctgacca
 15 2041 ggaatttggg gagctcagg acattgtctg gctttgggga ttcctccac atgctgcacg
 2101 cgcattcgc cccaggggc actgctgga agaticagga gcttggcgcg ccttcgcta
 2161 ctctacctg cttctgagtt gccagggag ccactggcag atgtccggcg gaagagaaga
 2221 gaacattgt tgaagaagc agccatgag agcgccctt cctgggactc gccctatcc
 2281 tctctctgt cccctctcg ggtgacgc taaaaggacc tatgtctca caccatgaa
 20 2341 accactagtt ctgtccccc aggaacctg gttgtgtgtg tgtgagtgt ttgacctct
 2401 ccatccctg gtctctctct tccctccg aggcacagag agacaggcca ggalccacgt
 2461 gccattgtg gaggcagaga aaagagaaag tgtttatat acggtactta tttatatcc
 2521 cttttaait agaaattaga acagttaait taattaaga gtgggtttt tticagtat
 2581 tcttggttaa tattttaait caactatita tgatgtat cttttgctct ctctgctct
 25 2641 ctatttga cgggttttg tatataaaat tcatgttcc aatctctc tccctgacg
 2701 gtgacagtca ctagctatc ttgaacagat atttaaitt gctaacactc agctctgcc
 2761 tcccgatcc cctggctccc cagcacacat tcccttgaag gagggttca atatacatc
 2821 acatacata tatatatgg gcaacttga ttigtgtga tatatatata tatatgtta
 2881 tglatatatg tgatccgaa aaaataaca tcgctattct gttttata tgttcaaac
 2941 aaacaagaa aaatagagaa ttctacatac taaatctc tccctttta attttaatat
 30 3001 ttgttatcat ttatttattg gtgctactgt ttatccgtaa taatttggg gaaaagatat
 3061 taacatcacg tctttgtc tagtgagtt ttccagata ttccgtagta catatttatt
 3121 ttttaacaac gacaaagaaa tacagatata tcttaaaaaa aaaaaa (SEQ ID NO: 12)

35 General Target Regions:

(1) 5' Untranslated Region - nts 1 - 701

(2) 3' Untranslated Region - nts 1275 - 3166

Initial Specific Target Motifs:

- 5 (1) Internal Ribosome Entry Site (IRES) in 5' untranslated region nts 513 -704
 5'CCGGGCUCAUGGACGGGUGAGGCGGCGUGUGCGCAGACAGU
 GCUCCAGCGCGCGCGCUCCCCAGCCUGGCCCGCCUCGGGCCG
 GGAGGAAGAGUAGCUCGCCGAGGCGCCGAGGAGAGCGGGCCGC
 CCCACAGCCCGAGCCGGAGAGGGACGCGAGCCGCGCGCCCCGGU
 10 CGGGCCUCCGAAACCAUGAACUUUCUGUCUCUUGGGUGCAUU
 GGAGCCUUGCCUUGCUGCUUACCUCCACCAUG 3' (SEQ ID NO:
 13)
- (2) Group III AU-Rich Element (ARE) Cluster in 3' untranslated region
 5' NAUUUAUUUAUUUAN 3' (SEQ ID NO: 10)

15

6.6. Human Immunodeficiency Virus I ("HIV-1")

GenBank Accession # NC_001802:

- 1 ggtctctctg gttagaccag atctgagcct gggagctctc tggctaacta gggaaccac
 61 tgcctaagcc tcaataaagc ttgccttgag tctctcaagt agtgtgtgoc cgtctgtgtt
 121 gtgactctgg taactagaga tccttcagac ccitttagtc agtgttgaaa atctctagca
 181 gtggcgcccg aacaggggacc tgaagcgaa agggaaacca gaggagctct ctgcagcgag
 241 gactcggctt gctgaagcgc gcacggcaag aggcgagggg cggcgactgg tgaatagcc
 301 aaaaattttg actagcggag gctagaagga gagagatggg tgcgagagcg tcaatattaa
 361 gcgggggaga attagatcga tgggaaaaaa ttcggttaag gccaggggga aagaaaaaat
 421 ataatataa acatatagta tgggaagca gggagctaga acgattcgca gttatccctg
 481 gctgtttaga aacatcagaa ggcgtgtagc aaactctgg acagctacaa ccatcccttc
 541 agacagatc agaagaactt agatcaatt ataatacagt agcaaccctc taitgtgtgc
 601 atcaaggat agagataaaa gacaccaagg aagcittaga caagatagag gaagagcaaa
 661 acaanagtaa gaaaaaagca cagcaagcag cagctgacac agacacagc aatcaggta
 721 gccaaatta cctatagtg cagaacatcc aggggcaaat ggtacatcag gccatatcac
 781 ctgaacttt aatgcattg gtaaaagtag tagaagaaa gctttcagc ccagaagtga
 841 taccatgtt ticagcatla tcagaaggag ccaccccaca agatttaac accatgctaa
 901 acacatggg gggacatcaa gcagccatgc aaatgttaaa agagaccatc aalgaggagg
 961 ctgcagaatg ggatagagtg catccagtgc atgcagggcc tattgcacca ggccagatga
 1021 gagaaccaag gggaagtgc atagcaggaa ctactagtac ccttcaggaa caaataggat
 1081 ggatgacaaa taatccacct atccagtag gagaaattta taaagatgg ataactctgg

1141 gattaaataa aatagtaaga atgtatagcc ctaccagcat tctggacata agacaaggac
 1201 caaaggaacc ctftagagac tatgtagacc ggftctataa aactctaaga gccgagcaag
 1261 cttcacagga ggtaaaaaat tggatgacag aaaccttgtt ggtccaaaat gcgaaccag
 5 1321 attgtaagac tatttataaa gcattgggac cagcggctac actagaaga atgatgacag
 1381 catgtcagg agtaggagga cccggccata aggcgaagat ttggctgaa gcaatgagcc
 1441 aagtaacaaa ttcagctacc ataattgagc agagaggcaa tttaggaac caaagaaaga
 1501 ttgttaagtg ttcaattgt ggcaaaagag gcacacagc cagaaattgc agggccoccta
 1561 ggaaaaaggg ctgttgnaa ttgtgaaagg aaggacacca aatgaagat tgtactgaga
 1621 gcagagctaa ttitttaggg aagatctggc ctctctacaa ggggaaggcca gggaatttc
 1681 ttcagagcag accagagcca acagcccac cagaagagag cttcaggtc ggggttagaga
 1741 caacaatcc cctcagaag caggagccga tagacaagga actgtatct ttaattccc
 1801 tcaggtcact cttggcaac gacccctct cacaataag ataggggggc aactaaagga
 1861 agctctatta gatacaggag cagatgatac agtattaga gaaatgagtt tgcagggaag
 1921 atggaaacca aaaatgatal ggggaattgg aggtttatc aaagtaagac agtatgatac
 15 1981 gatactcata gaaatctgtg gacataaagc tatagttaca gtattagtag gacctacac
 2041 tgtcaacata attggaagaa atctgttgac tcagatttgt tgcacittaa atttcccat
 2101 tagccctatt gagactgtac cagtaaaatt aaagccagga atggatggcc caaaagttaa
 2161 acaatggcca ttgacagaag aaaaaataa agcattagta gaaatttga cagagatgga
 2221 aaaggaagg gaaatttcaa aaattgggccc tgaataacca tacaatactc cagtattgc
 2281 cataagaaa aaagacagta ctaaaggag aaaattagta gatttcagag aactaataa
 2341 gagaactcaa gactctggg aagtcaatt aggaatacca catccgcag ggttaaaaaa
 2401 gaaaaaatca gtaacagtac tggatgtggg tgaatcatat tticagttc ccttagatga
 2461 agacttcagg aagtatactg catttaccat acctagtata aacaatgaga caccagggat
 2521 tagatatcag tacaatgtc ttccacagg atggaaagga tcaccagcaa ttttccaaag
 2581 tagcatgaca aaaatcttag agccitttag aaacaaaaat ccagacatag ttatctatca
 2641 atacatggat gattgtatg taggatctga cttagaata gggcgacata gaacaaaaat
 2701 agaggagctg agacaacatc tttgtagggtg gggtactacc acaccagaca aaaaacatca
 2761 gaaagaacct ccattccitt ggtggggtta tgaactccat cctgataaat ggacagtaca
 2821 gccatatgtg ctgccagaaa aagacagctg gactgtcaat gacatacaga agttagtggg
 2881 gaaattgaa ttggcaagtc agatttacc agggattaaa gtaaggcaat tatgtaaact
 2941 ccttagagga accaaagcac taacagaagt aataccacta acagaagaag cagagctaga
 3001 actggcagaa aacagagaga ttctaaaga accagttacat ggagtgtatt atgacccatc
 3061 aaaagactta atagcagaaa tacagaagca ggggcaaggc caatggacat atcaaattta
 3121 tcaagagcca tttaaaatc tgaaaacagg aaaatatgca agaattgggg gtgcccacac
 3181 taatgatgta aaacaattaa cagaggcagt gcaaaaaata accacagaaa gcatagtaat

3241 atggggaaag actcctaataa ttaaactgcc catacaaaaag gaaacatggg aaacatggtg
 3301 gacagagtat tggcaagcca cctggatfcc tgagtgggag ttgttaata cccctccctt
 3361 agtgaatta tggtagcagt tagagaaaga acccatagta ggagcagaaa ccttctatgt
 5 3421 agatggggca gctaacaggg agactaaatt aggaaaagca ggatatgtta ctaatagagg
 3481 aagacaaaaa gtgtcacccc taactgacac aacaatcag aagactgagt tacaagcaat
 3541 ttatctagct ttgcaggatt cgggattaga agtaaacata gtaacagact cacaatatgc
 3601 attaggaatc attcaagcac aaccagatca aagtgaatca gagttagtca atcaataat
 3661 agagcagtta ataaaaaagg aaaagggtcta tctggcatgg gtaccagcac acaaaaggaat
 10 3721 tggaggaaat gaacaagtag ataaattagt cagtgcctga atcaggaaag tactattttt
 3781 agatgggaata gataaggccc aagatgaaca tgagaataat cacagttaft ggagagcaat
 3841 ggctagtgtt tttaacctgc cacctgtagt agcaaaaaga atagtagcca gctgtgataa
 3901 atgtcagcta aaaggagaag ccatgcatgg acaagtagac tgtagtccag gaatatggca
 3961 actagattgt acacatttag aaggaaaagt tatccttgta gcagticatg tagccagtgg
 15 4021 atatatagaa gcagaagtta ttccagcaga aacagggcag gaaacagcat attttctttt
 4081 aaaattagca ggaagatggc cagtaaaac aatacatact gacaatggca gcaatttacc
 4141 cgtgtctacg gttagggccg cctgttggcg ggcgggaatc aagcaggaat ttggaattcc
 4201 ctacaatccc caaagtcaag gagttagtaga atctatgaat aaagaattaa agaaaattat
 4261 aggcacagga agagatcagg ctgaacatct taagacagca gtacaaattg cagttattcat
 20 4321 ccacaatttt aaaagaaaag gggggattgg ggggtacagt gcaggggaaa gaatagtaga
 4381 cataatagca acagacatac aaactaaaga attacaaaaa caaattacaa aaattcaaaa
 4441 ttttcgggtt tattacaggg acagcagaaa tccactttgg aaaggaccag caaagctcct
 4501 ctggaaaagg gaaggggcag tagtaataca agataatagt gacataaaaag tagtgccaag
 4561 aagaaaaagca aagatcatta gggattatgg aaaacagatg gcaggtgatg attgtgtggc
 25 4621 aagtagacag gatgaggatt agaacatgga aaagtttagt aaaacaccat atgtatgttt
 4681 cagggaaaagc taggggatgg ttftatagac atcactatga aagccctcat ccaagaataa
 4741 gttcagaagt acacatcoca ctaggggatg ctagattggt aataacaaca tattggggtc
 4801 tgcatacagg agaaaagagac tggcatttgg gtcaggggagt ctccatagaa tggaggaaaa
 4861 agagatatag cacacaagta gacctgaac tagcagacca actaatcat ctgtattact
 30 4921 ttgactgttt ttgactctt gctataagaa aggccttatt aggacacata gttagcccta
 4981 ggtgtgaata tcaagcagga cataacaagg taggatctct acaactcttg gcactagcag
 5041 cattataaac accaaaaaag ataaagccac ctttgcctag tgttacgaaa ctgacagagg
 5101 atagatggaa caagggccag aagaccgaag gccacagagg gagccacaca atgaatggac
 5161 actagagctt tttagggagc ttaagaatga agctgttaga caftttccta ggtattggct
 35 5221 ccatggctta gggcaacata tctatgaac ttatggggat acttgggcag gagtgggaagc
 5281 cataataaga attctgcaac aactcctgtt tatccatttt cagaattggg tgtcgacata

5341 gcgaatagg cgttactga cagaggagag caagaaatgg aggcagtaga tcctagacta
 5401 gagcccttga agcatccagg aagtcagcct aaaactgct gtaccaatlg ctatltgtaa
 5461 aagtgttgc ttcattgcc agttgttgc ataacaaag ccttaggcct ctccatggc
 5 5521 aggaagaagc ggagacagcg acgaagagct calcagaaca gtcagactca tcaagctct
 5581 ctatcaaac agtaagtagt acatgtaatg caacctatac caatagtagc aatagtagca
 5641 ttagtagtag caataataat agcaatagtt gtgtgttcca tagtaatcat agaatatagg
 5701 aaaatattaa gacaaagaaa aatagacaggg ttaattgata gactaataga aagagcagaa
 5761 gacatgtggc atgagagtga agggagaata tcagcacttg tggagatggg gtgtggagatg
 10 5821 gggcaccatg ctccctggga tgttgaatgat ctgtagtgc acagaaaaat gtgtgggtcac
 5881 agtctattat ggggtacctg tgtggaaggga agcaaccacc actctatitf gtgcatcaga
 5941 tgcataagca tatgatacag aggtacataa tgtttgggcc acacatgcct gtgtaccac
 6001 agaccccaac ccacaagaag tagtatgggt aaatgtgaca gaaaatttta acatgtggaa
 6061 aaatgacatg gtagaacaga tgcattaggga tataatcagt ttatgggac aaagcctaaa
 15 6121 gccatgtgta aaatlaaccc cactctgtgt tagtttaag tgcactgatt tgaagaatga
 6181 tactaatacc aatagtagta gcggggaagt gataatggag aaaggagaga taaaaactg
 6241 ctcttcaat alcagacaa gcataagagg taagggtcag aaagaatatg catttttta
 6301 taaactgtat ataataccaa tagataatga tactaccagc tataagtga caagttgtta
 6361 cacctcagtc attacacagg cctgtccaaa ggtatccttt gagccaatic ccatacatta
 20 6421 ttgtcccccg gctgtgtttg cgattctaaa atgtaataat aagacgtica atggaaacagg
 6481 accatgtaca aatgtcagca cagtacaatg tacacatgga attaggccag tagtataac
 6541 tcaactgctg ttaaatggca gtctagcaga agaagaggta gtaattagat ctgtcaatt
 6601 caccgacaat gctaaaacca taatagtaca gctgaacaca tctgtagaaa ttaattgtac
 6661 aagaccceac aacaatacaa gaaaaagaat ccgtatccag agaggaccag ggagagcatt
 25 6721 tgttacaata gaaaaaatag gaaatagtag acaagcacat tgtaacatta gttagacaaa
 6781 atggataaac actttaaanc agatagctag caaataaga gaacaatttg gaaataataa
 6841 aacaataatc ttaagcaat cctcaggagg ggaccagaaa atgtaacgc acagtittaa
 6901 ttgtggaggg gaattttct actgtaatc aacacaactg tttatagta cttgttttaa
 6961 tagtacttgg agtactgaag ggtcaaataa cactigaagga agtgacacaa tcacctccc
 30 7021 atgcagaata aaacaatfa taaacatgtg gcagaaagta ggaaaagcaa tgtatgcccc
 7081 tccatcagt ggacaaatfa gatgttalc aaatattaca gggctgctat taacangaga
 7141 tgggtgtaat agcaacaatg agtccgagat ctacagacct ggaggaggag atatgaggga
 7201 caattggaga agtgaattat ataataataa agtagtaaaa attgaacct tagtgatgac
 7261 accaccaag gcaagagaa gtgtgtgtga gagagaaaaa agagcagtggt gaattaggagc
 35 7321 ttgttcctt gggttcttgg gacgacgagg aagcactatg ggcgcagcct caatgacgct
 7381 gacggtacag gccagacaat tatgtcttgg tatatgtcag cagcagaaca atttgtctgag

7441 ggctattgag gcgcaacagc atctgttgca actcacagtc tggggcatca agcagctcca
 7501 ggcaagaatc ctggctgtgg aaagatacct aaaggatcaa cagctcctgg ggaattgggg
 7561 ttgctctgga aaactcattt gcaccactgc tgtgccttgg aatgctagtt ggagtaataa
 5 7621 atctctgga cagatttggg atcacacgac ctggatggag tgggacagag aaattaacaa
 7681 ttacacaagc ttaatacact ccttaattga agaatcgcaa aaccagcaag aaaaagaatga
 7741 acaagaatga ttggaattag ataaatgggc aagtitttgg aattggttta acataacaaa
 7801 ttgctgtgg tatataaaat tatcataat gatagtagga ggcttggtag gttaagaat
 7861 agtttttgc tgaacttcta tagtgaatag agttaggcag ggatattcac cattatcgtt
 10 7921 tcagaccac ctcaccaacc cgaggggacc cgacaggccc gaagggaatag aagaagaagg
 7981 tggagagaga gacagagaca gatccattcg attagtgaac ggaacctgg cacttatctg
 8041 ggacgatctg cggagcctgt gcctcttcag ctaccaccgc ttgagagact tactctgat
 8101 tgaacgagg attgtggaac ttctgggacg caggggggtg gaagccctca aatatgttg
 8161 gaatctccta cagtattgga gtcagggaact aaagaatagt gctgttagct tgcataatgc
 15 8221 cacagccata gcagtagctg aggggacaga taggggtata gaagtagtac aaggagcttg
 8281 tagagctatt cgccacatac ctagaagaat aagacagggc ttggaagga ttttgcata
 8341 agatgggtgg caagtgtca aaaagtatgt tgattggatg gcctactgta agggaaagaa
 8401 tggagcagc tgagccagca gcagataggg tgggagcagc atctcgagac ctggaaaaaa
 8461 atggagcaat cacaagtagc aatacagcag ctaccaatgc tgcctgtgcc tggctagaag
 20 8521 cacaagagga ggaggagggt ggtttccag tcacacctca ggtaccttta agaccaatga
 8581 ctacaaggc agctgtagat cttagccact ttttaaaaga aaagggggga ctggaagggc
 8641 taactactc ccaagaaga caagatacc ttgatctgtg gatctaccac acacaaggct
 8701 actccctga ttacagaac tacacaccg ggccaggggt cagatatcca ctgacctttg
 8761 gatgtgcta caagctagta ccagttgagc cagataagat agaagaggcc aataaaggag
 25 8821 agaacaccag ctgtgtacac cctgtgagc tgcattggat ggaatgacccg gagagagaag
 8881 tgttagagtg gaggtttgac agccgcctag catttcatca cgtggccga gagctgcatc
 8941 cggagtactt caagaactgc tgacatcgag ctgtctaca gggacttcc gctggggact
 9001 tccaggggag gcgtggcctg ggcgggactg gggagtgagc agccctcaga tctgcata
 9061 aagcagctgc ttttgctg tactgggtct ctctggttag accagatcg agcctgggag
 9121 ctcttggt aactagggaa cccactgctt aagcctcaat aaagcttgc ttgagtgtt
 30 9181 c (SEQ ID NO: 14)

Initial Specific Target Motifs:

- (1) Trans-activation response region/Tat protein binding site - TAR RNA - nts 1
 - 60
 "Minimal" TAR RNA element

5' GGCAGAUUCUGAGCCUGGGAGCUCUCUGCC 3' (SEQ ID NO: 15)

(2) Gag/Pol Frameshifting Site - "Minimal" frameshifting element
5'

UUUUUUAGGGAAGAUCUGGCCUUCUACAAGGGAAGGCCAGG
GAAUUUUCUU 3' (SEQ ID NO: 16)

6.7. Hepatitis C Virus ("HCV" - Genotypes 1a & 1b)

GenBank Accession # NC_001433:

1 ttgggggcga cactccacca tagatcaact cctgtgagg aactactgtc ttcacgcaga
61 aagcgtctag ccatggcggtt agtatgagtg ttgtgcagcc tcaggaccc cccctcccg
121 gagagccata gtggtctgcg gaaccggtga gtacaccgga atggcaggga cgaccgggtc
181 cttcttgga tcaacccgct caatgcctgg agatttgggc gtgccccgc gagactgta
241 gcgagtagt gttgggtcgc gaaaggcctt gtgtactgc ctgatagggt gcttgcgagt
301 gccccgggag gtctctaga ccgtgcata tgagcacaaa tctaaacct caaagaaaa
361 ccaaacgtaa caccaaccgc cgccacagg acgttaagt cccggcggtt ggtcagatg
421 ttggtggagt ttacctgttg ccgcgcagg gcgccagggt ggtgtgctgc gcgcatagga
481 agacttccga gcggtcgcaa cctcgtggaa ggcgacaacc tatcccaag gctcgcggc
541 ccgagggtag gacctgggct cagccgggtt acccttgccc cctctatgcc aacgagggt
601 tgggggtggc aggatggctc ctgicacccc gtgctctcgc gcctagttag gcccacacag
661 accccggcg taggtgcgtt aatttggga aggtcatcga taacctata tgcggtctc
721 ccgacctcat ggggtacatt ccgttgttc gcgcgccctt agggggcgct gccaggccc
781 tggcacatgg tgicccgggt ctggaggacg gcgtgaacta tgcaacaggg aatctgccc
841 gttgctctt ctctatctc ctctagctt tctgtcttg ttgacatc ccagcttccg
901 ctacgaggt gcgcaactg tcgggatat accatgtac gaacgactgc tcaactcaa
961 gtattgtga tgaggcagcg gacatgatc tgcacaccc cgggtgctgt cctctgctc
1021 gggagagtaa ttctccgtt tgcctggtag cgtcatcc cagctcgcg gccaggaca
1081 gcagcatccc caccagaca atacgacgc acgtgattt gctcgttgg gcgctgctc
1141 tctgtccgc tatgtacgtt ggggatctct gcggatcgt ttctctgc tccacgtgt
1201 tcaactctc acctcgccg tatgagacgg tacaagatt caattgtca atctatccc
1261 gccacgtac aggtcaccgc atggcttggg atatgatgat gaactgtca cctacaacgg
1321 cctagtgggt atcgagcta ctccggtacc cacaagcgt cgtggacatg gtgcggggg
1381 ccactgggg tgctctagcg ggccttgctt actattccat ggtggggaaac tgggtaagg
1441 tcttgattg gatgctact ttgctggcg ttgacgggca caccacgtg acagggggaa
1501 gggtagctc cagcaccag agcctcgtgt cctggctctt acaaggcca tctagaaaa
1561 tcaactcgt gaacaccaac ggacgtggc acatcaacag gaccgtctg aattgcaatg

1621 actccctcca aactgggttc attgctgcgc tgtctacgc acacagggttc aacgcgtccg
 1681 ggtgccccaga gcgcattgct agctgccgcc ccaltcgtac gttcgtcag ggggtgggtc
 1741 ccatcactca tgatactcct gagagctcgg accagaggcc atattgctgg cactacgcgc
 5 1801 ctgcaccgtg cgggacgtg cctgcgtcgc aggtgtgtgg tccagtgat tgccttacc
 1861 cgagccctgt tgtagtgggg acgaccgac gtctggcgc tctactgat agctgggggg
 1921 agaattgagc agacgtgctg ctacttagca acacggggc gcctcaaggc aactgtgttg
 1981 ggtgcacgtg gatgaacagc actgggttca ccaagacgtg cggggggccct ccgtgcaaca
 2041 tctgggggggt cggcaacaac accttggctc gccccacgga ttgcttcgg aagcaccocg
 10 2101 aggcacactta cacaagtgt ggctcggggc cctggttgac acccaggtgc atggttgact
 2161 acccatacag gctctggcac taccctgca ctgttaactt taccgtctt aaggtcagga
 2221 tgtatgtggg gggcgtggag cacaggctca atgctgcat caattggact cgagagagc
 2281 gctgtgactt ggaggacagg gataggctcag aactcagccc gctgctgctg tctacaacag
 2341 agtggcagat actgccctgt tcttcacca cctaccggc celtccact ggcgtgatcc
 15 2401 atcttcaccg gaacatcgtg gacgtgcaat acctgtacgg tatagggtcg gcagtgtct
 2461 cctttgcaat caaatgggag tatatcctgt tgcctttct tctctggcg gacgcgcgcg
 2521 tctgtgcctg cttgtggatg atgctgcfta tagccacgc tgaggccacc ttagaagaac
 2581 tgggtgtcct caatggcgcg tctgtgcgc gagcgcattg cctctctcc ttctctgtgt
 2641 tctctgcgc cgctcgttac atcaaaagca gctgtgtccc tggggcggca tatctctct
 20 2701 atggcgtatg gccgtgtctc ctgctcttgc tggccttacc accacgagct tatgccatgg
 2761 accgagagat ggctgcatcg tgcggaggcg cggtttttt aggtcttgta ccttgacct
 2821 tgcaccata ctataagggt ttctcgcta ggctcatat gttgttaca tatittatca
 2881 ccagagccga ggcgacatg caagtgtggg tccccctct caatgttcgg gtaggcgcgcg
 2941 atgccatcat cctcttaca tgcgcggtcc atccagagct aatcttgac atcacnaac
 25 3001 tctgtctgc cactactcgt ccgtcatgg tctccaggc tggcataact agagtccgt
 3061 actttgtacg cgtcaggggc ctatccgtg catgatgt agtcgggaag gtcgtcggag
 3121 gccactatgt ccaaatggcc ttcatgaagc tggccgcgt gacaggtagc tagctatatg
 3181 accattctac tccatcggg gattggggcc accggggcct acgagacctt ggcgtggcag
 3241 tagagccctt cgtcttctc gatcgtgaga cttaactcat cactcgggg gcagacaccg
 30 3301 cggcgtgtgg ggacatcatc tgggtctac cagtctccg ccgaaggggg aaggagatac
 3361 tcttaggacc ggccgafagt ttgagagac aggggtggcg gctccttgc cctatcagg
 3421 cctattccca acaaacgcgg ggctgtctt gctgtatcat cactagctc acagctcggg
 3481 acagaagca ggtcgtatgg gagggtcagg tgcctccac cgcaacgcaa tcttctcgg
 3541 cgacctcgt caatggcgtg tttggaccg tctacatgg tgcgggctcg aagaccttgg
 35 3601 ccggcccgaa ggttccaatc acccaaatgt acaccaatgt agaccaggac ctgctcggct
 3661 ggccggcgcc ccccgggggc cgtctcatga caccgtgcac ctgcggcagc tcggacctt

3721 acttggtcac gaggcattgct gatgtcgttc cgggtgcgccg gcggggcgac agcaggggga
 3781 gctcgttttc cccagggccc atctctacc tgaaggctc ctcgggtgga ccactgcttt
 3841 gcccttcggg gcacgttcta ggcactctcc gggtcgtctg gtgcacccgg ggggttgca
 5 3901 aggcgggtga cttcataccc gttgagtcta tggaaactac catgcggctc cgggtcttca
 3961 cagacaactc atccccccg gccgtaccgc aaacattcca agtggcacat ttacagctc
 4021 ccactggcag cggcaagagc accaaagtgc cggctgcata tgcagcccaa gggtaagg
 4081 tgctcgtcct aaaccgctcc gttccgcca cattggctt tggagcgtat atgtccaagg
 4141 cacatggcat cgagcctaac atcagaactg gggtaaggac catcaccacg ggcggcccca
 10 4201 tcactgactc cactattgc aagttccttg ccgacggtag atgtccggg ggcgcctatg
 4261 acatcataat atgtgatgaa tgcactcaa ctgactcgac taccatttg ggcacggca
 4321 cagtcctgga tcaggcagag acggctggag cgcggctcgt cgtgtccgcc accgccacg
 4381 ctccgggata gateaccgtg ccacaccca acatcgagga agtggccctg tccaacatg
 4441 gagagattcc ctctatggc aaagccatcc ccattgaggc catcaagggg ggaaggcacc
 15 4501 tcattcttcg ccattccaag aagaagtgtg acgagctcgc cgaaagctg acaggcctcg
 4561 gactcaatgc tttagcgtat taccggggtc tcatgtgtc cgtataccg actagcggag
 4621 acgtcgttgt cgtggcaaca gacgtcttaa tgcagggtt taccggcgac tttagctag
 4681 tgcactgact caacacatgt gtacccaga cagtcattt cagcttggat cccacttca
 4741 ccattgagac gacaacgctg ccccaagacg cgggtgcgc tgcgcagcgg cgaggtagga
 20 4801 ctggcagggg caggagtggc atctacaggt ttgtactcc agagaacgg cctcaggca
 4861 tgttcgactc ctggctcgtg ttgtagtct atgacgcagg ctgcgttgg tatgactca
 4921 gcgccctga gacctcgttt aggttgcggg cttaacctaa tacaccaggg ttgccgtct
 4981 gccaggacca cctagagtic tgggagagcg tcttcacagg cctaccac atagatgcc
 5041 acttcttgc ccagaccaa caggcaggag acaactccc ctacctgta gcataccaag
 25 5101 ccacagtgtg cgccagggtc caggctccac ctccatctg ggaccaaagt tggagtgtc
 5161 tcatacgtct aaagcccaca ctgcatgggc caacccccct gctgtacagg ctaggagcgg
 5221 tcaaaatga ggtcactctc acacaccca taacaaata catcatgga tgcattcgg
 5281 ctgacctgga ggtcgtcact agcacctggg tgcctagagg cggagtctt ggcgtctgg
 5341 ccgctactg cctgacgaca ggcagcgtgg tcaattgagg caggatcatc ttgtccggga
 30 5401 ggccagctgt tatcccgac agggaggtcc tctacagga gttcgatgag atggaagagt
 5461 gtgttcaca cctcccttac atcagcaag gaatgcagct cgccgagcaa tcaaacaga
 5521 aggcgtcggg attgtgcaa acagccaca agcaagcgga ggcgtcgtct cccgtgtggg
 5581 agtccaagtg gcgagccctt gaggctctct ggccgaaca catgtggaac ttatcaagg
 5641 gatacagta ctggcaggc ctatccactc tgcctggaaa ccccgcgata gcatcattga
 35 5701 tggcttttac agcctctatc accagccgc tcaccacca aaataccctc ctgtttaaca
 5761 tcttggggg atgggtggct gcccaactcg ctccccacg cgtgtctcg gcttctggtg

5821 gcgccggcat tgcgggtgcg gccgttggca gcataggtct cgggaaggta cttgtggaca
 5881 tcttgccggg ctatggggcg ggggtggctg gcgcactcgt ggcccttaag gtcatgagcg
 5941 gcgagatgcc ctccactgag gatctggtta attactccc tgcacitct tctctggcg
 5 6001 ccttggttgt cggggctgtg tgcgcagcaa tactgcgtcg gcacgtggcg ccgggagagg
 6061 gggctgtgca gtggatgaac cggctgtag cgltgccttc cgggggtaac cagctctccc
 6121 ccacgcacta tgtcccagag agcgcagccg cggcgctgt tactcagatc ctctccagcc
 6181 ttaccatcac tcagtgtctg aagaggcttc atcagtggat taatggagc tgcctcacgc
 6241 ctgttccgg ctctgggcta aaggatgttt gggactggat atgcacggtg ttgagtgaat
 10 6301 tcaagacttg gctccagtcc aagctcctgc cgggttacc gggactccct ttctgtcat
 6361 gccaacggcg gtacaaggga gtctggcggg gggatggcat catgcaaac acctgcccatt
 6421 gtggagcaca gatcacggga catgicaaaa atggctccat gaggattgtt gggccaaaaa
 6481 cctgcagcaa cacgtggcat ggaacatcc ccataacgc atacaccagc ggcctctgca
 6541 cgccctcccc agcccggaac tattccaggg cgctgtggcg ggtgctgctt gaggagtacg
 15 6601 tggaggttac gcgggtgggg gatttccact acgtgacggg catgaccact gacaacgtga
 6661 aatgcccatg ccaggttcca gccctgaat ttctacgga ggtggatgga gtacgggttc
 6721 acaggtatgc tccagtgtgc aaacctctcc tacgagagga ggtcgtatc caggtcggcg
 6781 tcaaccagta cctggtcggg tcacagctcc catgtgagcc cgaaccggat gtgacagttc
 6841 tcaattccat gtcacccgac cctctcata ttacagcaga gacggccaag cgtaggcttg
 20 6901 ccagggggtc tccccctcc ttggccagct cttagctag ccagttgtct gcgccttct
 6961 tgaaggcgac atgtactacc catcatgact ccccggaacg tgacctcatc gaggccaacc
 7021 tctctgtgcg gcaggagatg ggcgggaaca tcacctgtgt ggagtcagaa aataaggttg
 7081 taatctgga ctcttcgat ccgattcggg cggtgaggga tgagaggaa atactcgtcc
 7141 gcggcgagat cctgcgaaaa ccaggaagt tccccccagc gttgccata tgggcacgcc
 25 7201 cggattacaa cctccactg cttaggtcct ggaaggacc ggactacgtc ccccgggttg
 7261 tacacgggtg cctttgcca tctaccaagg ccccccaat accacctcca cggagggaaga
 7321 ggacgggtgt cctgacagag tccaccgtgt cttctgcctt gcgggagctc gctactaaga
 7381 cctttggcag ctccgggtcg tggccgttg acagcgacac ggcgactggc cctcccgatc
 7441 aggcctccga cgacggcgac aaaggatccg acgttagtc tgaactctcc atgccccccc
 30 7501 tcgagggaga gccaggggac ccgacatca gcgacgggtc ttgtctacc gtgagcgggg
 7561 aagctggta ggaactcgtc tgcgtctcaa tgcctatac atggacaggt gccattgata
 7621 cgccatgcg tgcggaggag agcaagtgc ccataatcc gttgagcaac tctttgtc
 7681 gtcaccacag tatgtctac tcacaacat ctgcagcgc aagctgcgg cagaagaagg
 7741 taactttga cagactgcaa gtctggagc accactaccg ggcgtgtcic aaggagatga
 35 7801 aggcgaaggc gtcacagtt aaggctaggc ttctatctat agaggaggcc tgcnaactga
 7861 cgccccca ttcggccaaa tccaaatttg gctacggggc gaaggagctc cggagcctat

7921 ccagcagggc cgtcaaccac atccgctccg tgtgggagga cttgctggaa gacactgaaa
 7981 caccaattga taccaccatc atggcaaaaa atgagggttt ctgcgtccaa ccagagaaag
 8041 gagggccgcaa gccagctcgc cttatcgtat tcccaacgtt gggggtagct gtaigcgaga
 5 8101 agatggccct ttacgacgtg gtctocaccc ttctcaggc cgtgatgggc cctcctacg
 8161 gatccagta ctctctggg cagcgggtcg agttcctggt gaatactgg aaatacaaga
 8221 aatgccctat gggcttctca tatgacccc gctgcttga ctcaacgggc actgagaatg
 8281 acatccgtac tgaggaaatc alttaccat gtgtgactt gggcccccga gccagggcagg
 8341 ccataaggtc gctcacagag cggctttatg tcgggggtcc cctgactaat tcgaaggggc
 10 8401 agaactcggg ttatcgccgg tgccgcgcaa gtggcggtgt gacgactagc tgcgggcaaca
 8461 ccctcacatg ttacttgaag gccactcggg cctgtcgagc tgcaaaagctc caggactgca
 8521 cgatgctcgt gaacgggagc gacctgtcgc ttatcttga gagtgcggga acccaggagg
 8581 atggggcggc cctacgagcc ttacggaggc ctatgactag gtattccgoc cccccgggg
 8641 acccgcccca accagaatac gacttgagc tgataacgtc atgctctccc aatgtctcgg
 15 8701 tcgcgcagca tgcaccggc aaaagggtgt actacctac cctgacccc accaccccc
 8761 tcgcacgggc tgcgtggagc acagttagac acactccagt caactctgg ctaggcaata
 8821 tcatcatgta tgcgccacc ctatggcgca ggaatgattt gatgactcat ttcttctcta
 8881 tctctctagc tcaggagcaa ctgnaaaag cctggattg tcagatctac ggggcoctgt
 8941 actccattga gccactgac ctactcaga tcaatgaacg actccatggt cttagogcat
 20 9001 ttctactcca cagttactt ccaggtaga tcaataggtt ggcttcacg ctacaggaaac
 9061 ttgggttacc gcttttgcga gtctgagac atcgggcccag aagtgtccgc gctaagctac
 9121 tgtcccaggc ggggagggct gccacttgcg gcaagtacct cttaactgg gcagtaaaga
 9181 ccaagcttaa actcactcca atcccggctg cgtccagct agactgtcc ggctggctcg
 9241 tigtggtta caacggggga gacatatafc acagcctgtc tctgcccga cccggttgt
 25 9301 tcatgttgt cctactccta ctttctgtag gggtaggcat ctactgtc ccaaccggt
 9361 gaacggggag ctaaccactc caggccaata ggccattccc tttttttt ttc (SEQ ID NO: 17)

General Target Region:

5' Untranslated Region - nts 1 - 328 - Internal Ribosome Entry Site (IRES):

30 5'UUGGGGCGACACUCCACCAUAGAUCACUCCCCUGUGAGGAACUACUGUCU
 UCACGCAGAAAGCGUCUAGCCAUGGCGUUAGUAUGAGUGUUGUGCAGCCUC
 CAGGACCCCCCUCCCCGGAGAGCCAUGAGUGGUCUGCGGAACCGGUGAGUAC
 ACCGAAUUGCCAGGACGACCGGGUCCUUUCUUGGAUCAACCCGCUCAAUGC
 CUGGAGAUUUGGGCGUGCCCCCGCAGACUGCUAGCCGAGUAGUGUUGGGU
 35 CGCGAAAGGCCUUGGUACUGCCUGAUAGGGUGUCUUGCGAGUGCCCCGGG
 AGGUCUCGUAGACCGUGCAU3' (SEQ ID NO: 18)

Initial Specific Target Motifs:

- (1) Subdomain IIIc within HCV IRES - nts 213 - 226
 5'AUUUGGGCGUGCCC' (SEQ ID NO: 19)
- (2) Subdomain IIIId within HCV IRES - nts 241-267
 5'GCCGAGUAGUGUUGGGUCGCGAAAGGC' (SEQ ID NO: 20)

6.8. Ribonuclease P RNA ("RNaseP")

10 GenBank Accession #s

X15624 Homo sapiens RNaseP H1 RNA:

1 atggcgagg ggaagctcat cagtggggcc acgagctgag tgcgtctgt cactccactc
 61 ccatgtccct tgggaaggtc tgagactagg gccagaggcg gccctaacag ggtctctcct
 121 gagcttcagg gaggtgagtt ccagagaaac ggggctccgc gcgaggtcag actgggcagg
 15 181 agatgccgtg gacccggccc ttggggagg ggcccgccgg atgcctctt tgcggagat
 241 tgggaacagc tcacggccag cgaagtgagt tcaatggctg aggtgaggtt ccccgacagg
 301 gacctataa ccaattcag accactctc tcgcccatt (SEQ ID NO: 21)

U64885 Staphylococcus aureus RNaseP (rrnB) RNA:

1 gaggaagtc cgggctcaca cagtctaga tgatttagt gticgtgctt gatgaacaa
 20 61 taaatcaagg catataattg acggcaatga aatatcctaa gtcttcgat atggatagag
 121 taatttgaaa gtgccacagt gacgtagctt ttatagaaat ataaagggtg gaacgcggtg
 181 aaccctcga gtgagcaatc caaatttgtt aggagcactt gtttaacgga attcaacgtg
 241 taaacgagac acacttcgctg aatgaagtg gtgtagacag atggttatca cctgagtacc
 25 301 agtgtgacta gtgcacgtga tgagtacgat ggaacagaac gcggcttat (SEQ ID NO: 22)

M17569 Escherichia coli RNA component (M1 RNA) of ribonuclease P
 (rnpB) gene:

1 gaagctgacc agacagtcgc cgtctctgct tcgtctctt cggggggagac gggcggaggg
 30 61 gaggaagtc cgggctccat agggcagggt gccaggtaac gcctgggggg gaaaccacag
 121 accagtgcaa cagagagcaa accgccgatg gccgcgcgca gcgggatcag gtaagggtga
 181 aagggtgcgg taagagcgca ccgcgcggtt ggtaacagtc cgtggcacgg taaactccac
 241 ccggagcaag gccaaatagg ggttcataag gtacggcccg tactgaacc gggtaggctg
 301 ctgagccag tgagcgatg ctggcctaga tgaatgactg tccagacag aaccggctt
 35 361 atcggtcagt ttacct (SEQ ID NO: 23)

Z70692 *Mycobacterium tuberculosis* RNaseP (mpB) RNA:

1 ccaccgggtta cgaatctggc gaccatggcc ccacaatagg gccggggaga cccggcgctca
 61 gtggtggggc gcacgggtcag taactgtcgc gcaacacggg gttgactgac ggccaatata
 121 ggctccatag cgtcgccgcg ggatacagta aaggagcaat ctgtgacgga aaagacggcc
 181 gacgacgtct tcaaatgtc caaggacgag aaggtcgaat atgtgcagct ccggttctgt
 241 gacctgectg gcatcatgca gcactcacg attccggctt cggoccttga caagacgctg
 301 ttgacgacg gcttggocct tgacggctcg tgatctcgc gggttcagtc gatccacgaa
 361 tccgacatgt tgcctctcc cgaaccgcg acggcgcgca tcgaccggtt ccgcgcggcc
 421 aagacgctga atatacaatt ctltgtgcac gaccggttca ccctggagcc gtactccgc
 481 gaccgcgcga acatgcgccg caaggccgag aactacctga tcagcactgg catcgccgac
 541 accgcatact tcggcgccga ggccgagttc tacatttgc attcgtgtag ctgcactcg
 601 cgcgccaacg gctccttcta cgaagggtgac gccactcgg ggttggtgaa caccgcgcg
 661 gcgaccgagc ccgacggcag tcccaaccgg ggttacaagg tccgccaca gggcgggtat
 721 ttccagtgg cccccaacga ccaatcgtc gacctgcgcg acaagatgt gaccaacctg
 781 atcaactcgc gttctatct ggagaaggcg caccacgagg tgggcagcgg ccgacaggcc
 841 gagatcaact accagttcaa ttgcgtgctg cagccgcgcg acgacatgca gttgtacaag
 901 tacatcatca agaaccgcg ctggcagaa ggcgaaccgg tcacgttcat gcccaagcgc
 961 ctgttcggcg acaacgggtc cggcatgcac tgcctcatgt cgtgttgaa ggacggggcc
 1021 ccgctgatgt acgacgagac ggggttatgcc ggtctgtcgg acacggcccg tcaatcatc
 1081 ggcgcgctgt tacaccacgc gccgtcgtg ctggccttca ccaaccgcg ggtgaactcc
 1141 tacaagcggc tggttcccg ttacgaggcc ccgatcaacc tggcttatag ccagcgcaac
 1201 cgttcggcat gcgtgcgcat ccgatcacc ggcagcaacc cgaaggccaa gccgtcggag
 1261 ttccgaagcc ccgactcgtc ggccaaccg talctggcgt tctcggccat gctgatggca
 1321 ggctcggacg gttatcaaga caagatcgag ccgacggcgc ccgtcgacaa ggaatctac
 1381 gagctccgcg cgggaaggcg cgcgagtcac ccgacagctc cgaccagct gtcagatgtg
 1441 atcgaccgtc tcgagcgca ccacgaatac ctaccgaag gaggggtgtt cacaacgac
 1501 ctgatcgaga cgtggatcag tttaagcgc gaaaacgaga tcgagccggt caacatccg
 1561 ccgcatccct acgaattcgc gctgtactac gacgtttaa gactcttgc agtcggggtg
 1621 tagagggacg ggcgtgtcgt tgccaggggc ggcgtcgagg ttittcgat ggtgacggg
 1681 gccggcaacg gcgcgccgac caccgtcgcg aagagccgt ttaagaact tcaaggactg
 1741 ttacgcggcg tgccacaacc cgttggcaa tcatctccg accgcgagc ggggttgtct
 1801 tcacatgcgc cgaactcaa gccactcgt cgcggcgcg tctgtcgcg gccggttcag
 1861 gttaatgtgc ggggattcgt cgtcgggcg ggcgtccag ctgaccaacg gccgactcaa
 1921 ctccgaaca ctttgcgac taccgcttt gccgcggcg tcaccgtag gtatgttgc
 1981 aggaattccc caccgtcgtc gtttcgccg ccggccgca ccgcgaccgc attgagctgg

2041 cgccccgggtc ccggcagctg gtcgggtggc ttgcccgcga ccaacaccag cgcgttgcgg
 2101 gcccggggtgg cgttcagcca ggccctgacgg agcagctcca cgtcggctgc gggaaccaga
 2161 tcggcggccg cgtatcacat cagggtatgc agcgtcagg ttgtgtgcag ggccgggaacc
 5 2221 ttgtgtcgtat gctgtagctg cagcaactgc acggtcattc cgtatgcggc cagtccgccg
 2281 cggcccgatt ttgtgtgtgt gttggggctg gcaccgcgcg gcaaccgctc ggcactgcata
 2341 cggggccttga tgcggcgaat ctgcgcgacc gattcagcgg acacaccgtc gggcgggatac
 2401 cgcgttttgt cgaccatccg taggaatcgc tgacccaact cggcatcgcc ggcaaccgcg
 2461 ttgtgcgcgta gcaggccctg gatctccat ggctgtgcc actgctcgtat gtatgcggcg
 10 2521 taggaccacca gggctcggac cagcgggacc ttgcggccct cgggtcgcgaa attgcgcgtc
 2581 agctccagcg gcggatcgac gctgggtgic ccacgcagcg ccggaacccg ctgcgcgac
 2641 gatgtcgacc atttcaccgc ccgtgcacg tcgacgccgg tggccgcctc acagacgaac
 2701 atcactgtcg catccgaccc gtagcccaac tcggaccac ccagccgacc catgcccgat
 2761 accgcgatgg ccggccgggc gcgatcgtc tcgggaaggc tggcccgcat gatgacgtc
 15 2821 agcggcgccct gcagaccgc caccacacac gacgtcaacg cccgcacac ctggtgac
 2881 tcgacagcgc cgagcagctc gcgcgaacc atgcgggcca gctctgcag acgcagcgtg
 2941 cgcgcgcggc cgtatggccc ctccgggtcg gggtagcggc tcgccgaggc gatcagcgcc
 3001 cgagccacgg cgccgggctc ggtctcagc agcttcgggc ccgcaaggcc gctcctgtac
 3061 tgcgtgatga cccgcggcgc gcgcatcaac agatccggca catacgcca ggtaccaag
 20 3121 acatgcatga gccgcttggc caccgcgggc ttgtcccga gcgttgcgac gtaccagctt
 3181 tcggtggcca gcgcctact gacgcggcg taggccagca gtccgcctc gggatcgggg
 3241 gcatacgaca tccagtcag cagcctgggc agcagcaccg actgcaccg tccgcggcg
 3301 ccgctttgat tgaccaacgc cgacatgtgt ttcaacggc tctgcggtc ctctagccc
 3361 agcgcggcca gccggcgccc cgcggcctcc aacgtatgc cgtggcgcat ctccaacccg
 25 3421 gtcgggcca tcgattccag cagcggttga tagaagatt ttgtgtgtaa cttcgacacc
 3481 cgcacgttct gcttctttag ttctcccg agcacccgg ccgactcgt ttgcgcactg
 3541 gggccgatgt gggccgcgcg ccgaccagc cgcactgcct cctcgtctc gggatcgga
 3601 agcagggtgg tgcgtttag ccgctgaac tgcagtgcgt gctcagcag cctgaggaa
 3661 tcatacgacg cgtcatgttt gcgcgcgtcc tcacgccga tgtacgcgc ttgcccaac
 30 3721 gccgccaatg cgtccaccgt ggcgcgacc cgtaacgact cgtcgtacg ggcattgaac
 3781 agctgcagta gctgtacgc gaactcacg tcgcgaact cgcgcgtgc gagtttgagc
 3841 tcgcccggcg ggcacatggc gggcaccagc tgcctcaccc gccgcgcat ggcctgcac
 3901 tcgaccacaa agtcttcgc ctgcgaagct cgcacacca tcggcataa ggcggtcagg
 3961 taacgtctgc caagtccgc gtcgccaacg actgcgccgt ctttcagcaa cgcctgaac
 4021 tccaggtct ttgcccagcg ctggtagtag gcgatgtcgc actcagcgt acggaccagc
 35 4081 tccccttgc gccctccgg acgcaggcg gcgtccact cgaaaaagg cgccgaggcc

4141 accccgcatca tctcgtggc cagcgcgcgc ttgcgcggt cggagcgctc ggcaacgaat
 4201 atgacatcga cgtcgtgac glagtcagt tcgcgcgcac cgcactlgcc catcgcgatg
 4261 accgccaggc gcggtggcgg gtcctccgc cacacgtcg cctggccac gcgcagcgcc
 5 4321 gccgccagag cggcgtccgc ggctccgcc agcgctgcgg ccaccacggt gaatggcagc
 4381 accggttcgt cctgcaccgt cggggcagg tcgagagcgg ccagcattag cacgtagtgc
 4441 cggctactgg ttgcgaatgc gtgcacgagc gagccggca taccctccga ttctcgacg
 4501 cactcgacga acgaccgctg cagctgtgca tgggacggca gtgtgacct gcccgcgacg
 4561 aatttcagc actgcggatg ggcgaccagg tgatgccca acgccagcga cgaagccagc
 10 4621 accgagaaca gccgcccgcc cagactgcgt tcgcgcgaca gagcccggtt gactcgtcc
 4681 catccggtgt ctgattctc cgacagccgg atcaaggcgc gcagcggcg atcggcgctc
 4741 ggagcgcgtg acagcgacca cagcaggctg acgtgcgcct gatcctcgt ccgatccac
 4801 ccagctgag ccagacgctc accagcaggg gggtcaacta atccgagccg gcaacgctg
 4861 ggcaacttcg gccgtgcgt ggcgagttg gtcacacca cgacggtagc gcaagcgccg
 15 4921 tcggcgtcgc atcaaccggt agatctgggc tacagcgaca ggtagggtgc cagctcgtat
 4981 ggcgtgacgt ggtcgggta gtgcgccac tcgtgcgt tgttgccaa gaaaagta
 5041 aaaaagctct ccccaaggc ctccgcgacg agtcggagg cctccatgac gcgcagcgca
 5101 ctatcaaac tggacggcaa ttctcgta cccactgct ggctgtctc ggtgtgagg
 5161 tccatacgt tgtctcggc ctgcgggcc agcagtaac cttctctac accccgaat
 20 5221 cccgcggcca gcagcagcg gaatgtcaga tagggattgc acggcgaatc aggcgtgcgt
 5281 acttcgaccc gccgcgacga ggtctgtgc ggcgtgtaca tcggcaccgc cactaggcg
 5341 gatcgggttg cggccccca cgacgcggcc gtcggcgctt ccgcgccctg caccagccgc
 5401 ttgaagagt tgaccactg attgtgacc gcgtgatct cgcaagcgtg ctccagatc
 5461 ccggcgatga acgatttacc cactccgac agctgcagcg gatcatcagc gctgtgaa
 25 5521 gcgtlgacat caccctcga cagctcatg tgggtgtgca tcgccgagcc cgggtgtcgtg
 5581 ccgaatgct tggcatgaa cgacgccgg gcgccctt ccagcgcgac ttcttgatg
 5641 acgtagcga aggtcatcac gttgtgacc atcgacagag cgtcggcaaa ccgaggtgc
 5701 atctctgct ggcgggtgc gcctcglga tggctgaact ccaccgagat gccatgaat
 5761 tccagggcat cgtcgcgtg gcggcgaaag ttcaaggcgg agtcgtgac cgcttggtg
 30 5821 aaatagccgc cgtgtgcac cgggacgggc accgaccgt cctcggttc gggcttgagc
 5881 aggaagaact cgatttcgg atgcacgtag caggagaagc cgaatgtcc ggccttcgc
 5941 agctgcgcc gcaacacgtg ccgcgggtcc gccacgacg gcgagccgtc cggcatggtg
 6001 atgtcgaaca acatccgcgc tgagtgtgg tggccggaac tgttgccca gggcagcacc
 6061 tgaaggtgc acgggtccg gtgcgccacc gtaicggatt ccgagaccgc cgcaagacc
 35 6121 tcgatcgagg atccgtcga gccgatgct tctcgaagg gcgcctcgag ttgcgtggg
 6181 gcgatggcga ccgacttgag gaaaccgagc acgtctglga accacagccg gcgaagcgg

6241 atgtcgcgtt cttcagggt acgaagaac aattccttct gtcggfccaat acctcgaaca
 6301 gtatgcactg tctgttaaaa ccgtgttacc gatgcccgcc cagaagcgtt gcggggcgccg
 6361 ccgcgaaggcg agtgcgcggt gactcaggcg cgcgcaccgc agactcgtcg gcgcgaagg
 5 6421 cccgtcgaga aaatagtgca tcaccgcaga gtccacacac tggttgccat cgaacaccgc
 6481 agtgtgttgg gtgcccgcga aggtgatcag cgggtgcgcc agctggcgccg ccagggtctac
 6541 cccgactga tacggagtg cggggtcgtg ggtgtgtggac accacgacga ccttgcacgc
 6601 cccggccggc gccgcggggg cgggcgtcga cgttgcggcg accggccaca gcgcgcacgc
 6661 atcgcggggg gcggatccgg tgaactccc gtactaagg aacggcgga cctgacgcat
 10 6721 ccgttgcgtc gcggccaccc agggccgtgg atcggccggt gtggcgcat cgagccaccg
 6781 gaccgcgttg aacgcgtcct ggtcgttct gtatgcccgc tctgcatccc ggccgtcata
 6841 gtcgtcggca agcaccagca agtcgccggc gtcgtgcgcg cgtgcagcc ccagcagacc
 6901 actggtcagg tacttcacgc gctgagggt ctacagcgcg ttgatgtgtc ccgtgtcgc
 6961 gtcggcgtag ctacggccac gtggatccga cgtcttacc ggcttctga ccagggggtc
 15 7021 aaccaggcg tggtagcgtt tgaccacgtg gcgcgagtcg gtgccagag gcgaggccgg
 7081 cgagcggcg cagtcggcg cgtatcatt gaaacggtc tgaatcccg ccatttggct
 7141 gatctttcc tcgattggcg taacggctgg atcgatagcg ccgtcgagga ccatcgccc
 7201 cacatagta ccgaaccgtt ccagtgaa cgtggccaac tgggtccgt agctgtatcc
 7261 gaggtatgt atctgatct cactaacgc ttggcgaaac atgtccatg cccgtcgac
 20 7321 ggacgcgga ccgatatgg ccaagaagct gaagcccatc cgtcaacac agtctggcg
 7381 caactgccgg tagacctgtt cgactgggt gacaccggcc ggaactgtat cggccatccg
 7441 atcgcgccgg tacgcgtga actcggcgtc ggttcgcacac cgcaacgcag gggtcgagt
 7501 gccgacctt ctcgggtcga agcccaccag gtcgaagtgg cggagaatgt cgtgtcggc
 7561 gatcggggt gccatagcgg cgaccatgtc gaccggcgac gccccgggtc cccaggatt
 25 7621 gaccagcagt gctccgaatc gctgtccgt cgcggggacg cggatcaccg ccaacttgc
 7681 ttgttccca ccgggttgg cgtatgcga ggggacggac accgtcgcgc agcgtgcagt
 7741 gcgaatttc ctggtgtcgg cgtatgaatc gcgcagctg ttcaactct gttcgcgcc
 7801 cagcaccggc gcacccgggg ttggccggc gccgggttct tcatgtcgc cggccaacgg
 7861 gggcgctgct aggggcagtc cgccgagcga caaccggaag gacagcagc ccgagctaa
 30 7921 cgtgtcgcg cgcacatgg ccgcatcgt ctacccggcg aataccttg accgcgcgaa
 7981 atgacacac cttcgtttct tcgcccgcgt agcattggc gcgcgtggcg ggcgtgtgtc
 8041 cgcgattaa atacgcgcgc actactcgt caatgcagt gtcgocctgg aatacacgc
 8101 ttgtctgggt tcgtcgaag gtcagcaacg aaccgcgaag ctggtcgc accgtcgacc
 8161 cggccttga cgcgcgtcgc gggtcatggg ttggtgatac caccaccgtt ggcactaggc
 35 8221 cgggcgcga gacggcatgg ggtgacttg ttggtggcac cggccagaa gcgcaggtgc
 8281 ccagcgcgc atcaccgttg aactccgt agctcagaa cgtgtcgatc tcccgggcgc

8341 gggcggtcttc gtcgatgacc ttgtcgcgat cggtaaccgg gggctgacgc acgcaattga
 8401 tcgccaccgg cgcgtcaccg gaattgttgt agcggccgtg cgagtcccga cgcgatgaca
 8461 tgtcggccag agccagcagg gtgtctccgc gattgtcgac cagctccgac agcccgctgg
 5 8521 tcaagtgttg ccacagattc ggtgagtaca gcgcataat ggtgcccaag atggcgctgc
 8581 tataactcag cccgcgcgga tctctgtgic gcgcggccct gctgacctc ggggtgtccg
 8641 ggtcgaccaa cggatcgacc aggtctgtgt agacctcgac ggccttgccc ggtcgccgc
 8701 ccagcgggca gcccgcgttc ttggcgagcgt cggcgccata gttgtgaac gcgtctgga
 8761 agcccttggc ctggcgagc tccgcctcga tggatcggc attgggctgc acgcaccgt
 10 8821 cgagaatcat tgcccgcacc cgtctcgga attcctcggc atacgcggag ccgatccggg
 8881 tggctagaca gtatgccagg taggtcagct tgtcgtcgcc caacgccgcg cgaatggcat
 8941 ccaggtcctt ggcgacgttg accgtccga catgggccag aaagtcttg ccatcttgt
 9001 ccacacagcg accgagaat tgcctgtct cgttctgat gtgcgcaca cctcccgcg
 9061 tglagtcaac ctgcggctgc gccgcagcc ggtcgtgic ggcacggag ttgaccaga
 15 9121 tcgccggccg ggcgagcgc accccgcggg ggtcgaacc aaccagctgc aacttttgt
 9181 gcaccgctt cggcaatgtc tggagagcg ccaaggcgc ctgcatacc gattcgccg
 9241 gtccaccggc attatgacc agcgaaccga tctgtctcc cgtcgccga aagcgaalca
 9301 gcccagcgc gcgccagta ccatcgggc ggtcgtagtc gaccgttaca gcgagcttgc
 9361 cgcataacgc gccgccggg atctttactt gcgggttga cgcaccgcac ggtgtccact
 20 9421 ccaccgctg gccacgttc ggtccgccca tacgagcgcg tcccgcacc acgcggatgc
 9481 agcccacaag aaccaacgcc acggcgccga gcgcggccca gatcaacag atgcgcgcga
 9541 tctgtgcgc gcgagacag ctcatgccca caatgctgcc agagcagacc cgaatctgc
 9601 gccagcgccc accgtcgccc gactaacgg ccgtcgccag cagtcttgc atcgccgatg
 9661 gcgaactcgt cggccatccc ccatagtcg gtaacagat ccggcgaaga caccagccc
 25 9721 tcgaccggat ccggcaggg cgcgtcgccc tcggcggtgc acaactgcga catcaggttg
 9781 ggcctggcac ccgtccacg ccggcatgtt gcacttggc catcgccga ggcgatccc
 9841 cgatgccgt cacccttgc acgaacctat ctccacggc ggtcgccgc agcgacgcga
 9901 tgtggcgca gatctccgag agtctggccc gccgcggcg cgacggcaac ccgatgccgt
 9961 cgaagtgcg atcgatgta ggttaagggt tcagcgcaat gctggaagc ttittccgaa
 30 10021 accgcggcct cgccttgatc tggagtcaga acggtcgac cagccggta aagcgtaac
 10081 ccatctcga gcaaacatgc atggcgtgag tggacgttc cagacacgc aactgcgtc
 10141 caggccactg agccgtgca tgcgcgatgg tatgcgatg ggggccccg gcgcgtctga
 10201 ggggaagaag ttgcagactg tcagggtccg acgaaccgg ggcacctaac gggccacgag
 10261 gatcgacccg accaccatga ggcagagtga tgtctgagca gactatcat ggggccaata
 35 10321 ccccggagg ctccggggcg cggaccaaga tccgcacca ccactacag agatggaagg
 10381 ccgacggcca caagtggggc atgctgacgc cctacgacta ttgcagccc cggatcttcg

10441 acgagggccg catcccggcg ctgctggctg gtgattccgc ggccaacgtc gtgtacggct
 10501 acgacaccac cgtgccgacg tcatcgacg agctgatccc cttgggccgt ggctgtgtgtc
 10561 ggggtgcccc gcacgcactg gtcgtgccg acctgccgtt cggcagctac gagggggggc
 5 10621 ccaccgccgc gttggccgcc gccaccgggt tctcaaggc cggcggcgca catcggttca
 10681 agtcgcaggg cgggtgagcgg gtggccgagc aaatgcctg tctgaccgc gcggggcaicc
 10741 cgggtatggc acacatcggc ttacccccgc aaagcgtcaa cacttgggc ggcttcgggg
 10801 tgcagggccc cggcgacgcc gccgaacaaa ccatcgccga cgcgatcgcc gtgcgcgaag
 10861 ccggagcgtt tgcgctcgtg atggagatgg tgcgcccgga gttggccacc cagatcacgc
 10921 gcaagcttac cattccgacg gtcgggacg gcgctgggcc caactgcgac ggccagggtc
 10981 tggatggca ggacatggcc ggggtcagcg gcgccaagc gcgccgttc gtcaaacggt
 11041 atgcgcatgt cgggtgtgaa ctacgccgtg ctgcaatgca atacgccaa gaggtggccg
 11101 gcgggggtatt ccccgctgac gaacacagtt tctgaccaag ccgaatcage ccgatgcgcg
 11161 gggatctcgg tggcgccctg gatccgctcg acgcgggatt gccggcgcg acgcgccagc
 15 11221 gggaccacat ggctgcgctg tgcgggttg agccgggggt gagccagac aticgatgtg
 11281 cccaacacca tccgccacag cccaattgat gtggcactct atgcatgct atccccgac
 11341 aaccaccacc gcggcgacgc atcatgaccg gaggcgaaga tgcagtaga ggcgcccgga
 11401 ccagcgccgc atctggaggt cagagcgcaag ttcgacgtga tcgagtgcac ggtgtcgcg
 11461 tctgtcaggg gcatcgccgc ggtgtgtcgc gtgcagcagt cgcgcacca gcagctgcac
 11521 cgggtgtact tcgacacacc gtgcacgac ctggcgcgca accagatcac ctgtggcgcc
 20 11581 cgcaccggcg gcgcgcagcg cggctggcat ctgaagctgc cggccggacc cgaacaagcg
 11641 accgagatgc gagcaccgct gtccgcatca ggcgacgctg tgcggccga gttgttgat
 11701 gttgtgtctg cgaatgtccg cagaccggcg gttcagccgg tgcgcggat cagcactcac
 11761 cgcgaagacc agatcctgta cggcgccggg ggcgacgcgc tggcggaatt ctgcaacgac
 25 11821 gactgcaccg catggtcggc cggggcatlc cacgccgctg tgcagcgga caacggccct
 11881 gccgaacagc agtggcgcg atgggaactg gaactgtgta ccacgagatg gaccgcgat
 11941 accaagctac tggaccggct agccaaccgg ctgctcgatg ccggtgcgc acctgccgc
 12001 cagcgctcca aactggcgcg ggtgtctggt gcgacctct ccggtgagct gccaacggc
 12061 ccgcagccgc cggcggtacc agtacaccgc gcggtgtccg agcaaglcga gcagctgtcg
 30 12121 ctgtgggcat ggccggtgcg ggccgacgcc tatgacgcc tgcaccagat gcgagtgcag
 12181 accgcgaaga tccgagctt gctgacggat tccaggagt cgttggcct gaaggaaagt
 12241 gctgtggata tcatgaact gcgtgagctg gccgatgtcc tggcgctagc ccgggacgcc
 12301 gaggtactcg gtgaccgcta ccagcgcgaa ctggacgcgc tggcgccgga cgtgtacgc
 12361 gccgggggtc gcgagcgctt ggtagacgg gcgcggcggc gataccagac cgggtcgcg
 35 12421 cgaactacta tgcattgcg gtgcagcgg tacttccgtc tgcgcagc tctagacgcg
 12481 cttgtgtccg aacgcgcca tgccactct ggggaggaat cggcaccggt aaccatcat

12541 ggcgcctacc ggcgagtcgc caaagccgca aaagccgcaa agaccgcccgc cgaccaggcg
 12601 ggcgaccacc accgcgcagca ggcatlgcac ctgatccga agcgcgcgaa gcgattacgc
 12661 tacaccgcgc cgctactcgc ggcggacaat gtgtcacaag aagccaaggt catccagacg
 5 12721 ttgctaggcg atcatcaaga cagcgtggtc agccgggaac atctgatcca gcaggccata
 12781 gccgcgaaca ccgcgcgcga ggacaccttc acctacggtc tgcctacca acaggaagcc
 12841 gacttggccg agcgcctccg ggagcagctt gaagccgcgc tgcgcaaat cgacaaggcg
 12901 gtccgcaaa gacgggatlg agcccgcagc gggcggacga gttggcctgt aagccggatt
 12961 ctgttcgcgc ccgccacagc caagctaagc gcggcacgcg ggcgaccalc catctggaca
 10 13021 caccgttacc gggcgctcgc agcggcctac ccgcaggctc gggcgagcaa cctcaagcg
 13081 cctgcgcgcg cgcacttgc gtgcggcctt ctggccttg ctccgggtgg ggttgccta
 13141 gccaccccg gcacccggaa tgcgtggcgc ctctaccgc accgtttcac cctgccacc
 13201 acgaggatgg cgggtctgtt tctgtggcac ttccgcga gtcactcgg atfocggtta
 13261 gcaatcacc tgcctgtgga agtcgggact ttctcgact cgacgtgaa cctcgtgaat
 15 13321 ccaacaaagc cctacgcgag ccgcggccgc ccagccaact catccgcagc gaccacgta
 13381 ccccgctggc cgggtgcgcg gccagtgtga ccgctggacg acacgctag tggacagcc
 13441 gatccggcgc gcagtcctta tctgtgactg gtgacacggt gggacaacgc cgtcgactcc
 13501 ggcgactcgc accgcctcgc tgcgcaggtc agcgagtagc gtggcgcaat gctactcgcg
 13561 ctgatcacc ccgcgcgagc gccccggctg ccgaagctgt accgcgacga cgcctgttt
 20 13621 cgtcgcagcg tcgatatggc atccaagcg tacggcgccg ggcgctatcg atatttcat
 13681 gccccctatc ccgagtgatc gagcgtctca agcaggcgct gtatcccaa ctgtccga
 13741 tagcgcgcaa ctgtggggcc aaactgggccc gggaggcgcc cttgccagac agccttgatg
 13801 actggttggc gagctgtcat gccgcggccc aaaccgac cagcgcgtg atgttgaagt
 13861 accgcaccaa cgactggaac gccctacac aggatctcta ccgcgagttg gtgttccgc
 25 13921 tgcaggtggt gatcaacctg agcgatccgg aaaccgacta caccggcgcg gatttctgc
 13981 ttgtgaaca gcgcctcgc gcccaatccc ggggtaccgc aatgcaact ccgcaggagc
 14041 atggttatgt gttcacgacc cgtgatcgc cgttcgggac tagccgtggc tggtcggat
 14101 ctccagtgcg ccatgggctt tcgactatc gtccggcga acgctatgcc atggcgctga
 14161 tctttcaga ccgagcctga ttgcagcca tctatagata gctgtctga ttcaacaatc
 30 14221 gcaccagca tgcctcatcg gcgtagaact ccgcgatgct cagcgatgc agatcaagt
 14281 gcaaccgata taggacgcc gccccggcat ccaacgcagc ccgcaacaac attttgatc
 14341 gctgtacatg tgacaccac agcacgtcg cgccttcgta gccacgatg atccgatcac
 14401 gtccccgcgc aaccgcgcg agcacgtcgt cgaagcttcc ccaccggcg ggcgtgatgc
 14461 tgggtgctcg cagccagcga cgtgtcagct cgggatcgcg ttctgcgcgc tcccggaacg
 35 14521 tcagccctc ccaggcgcg aagtcggtct cgcaccagtc gtcacgagc accacgtcca
 14581 gggccaggcg tctggcgcg gtcaccgcg tctcgtgaag ccgctgtagc ggcgaggaga

14641 ccaccgcgagc gatcccgccg cgccgcgccca gatacccgcg cgccgcacca acctggcgcc
 14701 accccacctc gttcaacccc gggttgcgcg gccccgaata gcggcggtgc tccgacagct
 14761 ccgtctgccc gtggcgcaac aaaagtagtc ggggtgggtgt accgcggcgcg ccggtccagc
 5 14821 cgggagatgt cgggtgactcg gtgcgaacga ttttgcaggc atccgcattc gccgcagccg
 14881 attgcgcggc ggctgccatc gcgtcatigc ccaaccggtc tgcatactgc ttccggggac
 14941 gcggaaacca ctctgatttg atctcgcaa actgggagcg caacgcctga gcctggacat
 15001 agagcttcag cagatccggg tgcctgacct tccaccgcc ccgacatctgc tccaccaaca
 15061 gcttggagtc catcagcacc gcggcctcgg tggcacctag ttccacggcg tgcaccaaac
 15121 cggctatcag gcccggggat tccggcgact gtgtcgtcgc ccggccgfat gccgtcgtgg
 10 15181 actcgccag caccgtggag tgatcgccgg tccacaccac cgcgcgcgat ccggcccggtc
 15241 cgggattgcc ccgcgatccg ccgtcggctt cgaatgacaa ttcaactct caaatcttc
 15301 gagccgcaac aagatcgctc cgcattccgg gcagcgcaac acttafctct ccggccgcgc
 15361 cgagatctgg gccagctcgc cgcggccgat ctgcattccg caggcaccac atcgatgacc
 15 15421 ttgcaaccg ccggccctgc gccgcctcc ggcccgctgt ctttctaga gccccgcaag
 15481 ctccggatca agtgtgcgcg tcagcatgtc gcgttgcgat gaattgttgt gccgggcttg
 15541 gtccatttgc gcaagtgcct cgtccaaagc ctgctggcg gcggccaggt ccggccgcga
 15601 ccgttggagc gcccgcgact cggcggtctg ttgagcctgc agctcctgc gccgttcag
 15661 caccctcagc agggcatctt ccaaactggc ttgacggcgt tgcaagctgt ccagctctgt
 20 15721 ctgcagatca gccaatgtct tggcgctcgt tgcaccgaa gtgagcaac accgttccg
 15781 gtgcaccagc ttacgcaccg catcgatctc cgactcaaaa cgcgacacct gccgcacca
 15841 gtctccgcc gcgattgca gggccgcat cctgtcgttg gcggcggtgt gcctggcctg
 15901 caccctgtgg taagccgcc gcctggcgag atgggtagcc cgaatcgcca tccgggtcag
 15961 ctacgacatc agcttcgcca attccagtag cgaccgttg tgtgccact ccgcttcat
 25 16021 gcctgatctc tccagtttc gtgatcgagg ttccagggt cgggtcgagt ggtgcacaca
 16081 cgcaccgcga gcgacgcgcc gaattgagac cgaacactt ccggccgctg gcgcaccac
 16141 ggggaattgc ttgccaatg cgcagctgc atcaggccca ctgccaagc tccgcaatgc
 16201 tctgtggctg gatgatgtcg cagatcgcc gtaactgac ctgcaactc ccggcgccgc
 16261 accgtggcaa gcaacgagtc ccggcgccg ccgcaagacc cgaccgcga caccgacag
 30 16321 tgggataccc cggcgcgcg cacaccggtc gcatgcgcg gcaaccggc ctccagacg
 16381 gcaacaagg tgcgacgcg ttcggtttt ggcagtctgc caalccggc taaccgctg
 16441 ccgaccggcg gtgtaccag cgcgaagatg tgaatgcg gctcctgta aggggtgcgcg
 16501 gcgcgcatcg ccgccaacac ctggcgcg gcctgtgcgg gtgcgacga ctccagccg
 16561 tctcggcca ccgttgcag ggtaccgacg ctgcctatgg ccggcgacgc ccgtctgtc
 35 16621 gccaggaact gcccggtacc cgcgacactc cagctgcagt gcgagtgtg gccgatatg
 16681 ccggcaccgg cctcaagac gcctgccgc accgcctctg agttctcgc cggcacatag

16741 atgaccact tgtcgagatc ggccgctccg ggccaccgggt cgagaacggc gtcgacggtc
 16801 agaccaacac cgtgtgcccag cgcgtcggac acaccggcg acgcccagtc ggcgttggtg
 16861 tgcgcggtaa acaacgagcg accggtccgg atcaggcggf gcaccagcac accctttggc
 5 16921 gtgttggccg cgaccgtatc gacccacgc agtaacaacg ggtgtgtcac caatagcagt
 16981 ccggcctggg gaacctgggc caaccaccgc ggcgtcgcgt ccaccgcaac ggtcacggaa
 17041 tccaccacgt cgtcggggtc gccgcacacc agaaccaccg aatccacaga ctgggcaagg
 17101 cgcggcgggt aggcctggtc cagcacgtcg atgacatcg ccaggccgac atctatcggc
 17161 gtctctcacg ctttcccac tcggcgatcg ccgccaccag caggggccac tccggcgca
 10 17221 ccgcccggcg caggtaccgc gcgtccaggc cgacgaaggt gtcaccggcg cgcaccgaa
 17281 ttctttgtct ctgcaaatag ttctgaatc cgtcagcatc ggcatgttg aacagtacga
 17341 aagggggccg accatcgacc acctcggcac ccaccgatct cagtccggcc accatctccg
 17401 cgcgcagcgc cgtcaaccgc accgatcgg ctgcggcagc ggccagccgc cggggggcg
 17461 agcaagcagc gatggccgtc agttgcaatg ttccaacgg ccagtgcgtc cgtgtcacgg
 15 17521 tcaaccgagc cagcacgtct ggcgagccga gcgcgtagcc caccgccaat ccggccagcg
 17581 accacgtttt cgtcaagcta cggagcacca gcacatcggg cagcgagtca tcggccaacg
 17641 attgcgcgtc gccgggaacc caatcagcga acgctcgtc gaccaccagg atgcgtcccg
 17701 cggcgcttaa ctgagcagc tgcctcgga ggtgcagcac cgagtgggg ttgctcggat
 17761 taccacgac gacaagctcg gcgtcgtcag gcacgtgcgc ggtgtccagc acgaacggcg
 20 17821 gctttaggac aacatgtgc gccgtgattc cggcagcgt caaggctatg gccgcctcgg
 17881 tgaacgggg cagcagatt gctgcccgca ccggaactag gttgtgcagc aatgcgaatc
 17941 cctccgccgc ccgacgagc gggagcactt cgtcacgggt tctgcatga cgttcagcga
 18001 ccgcgtcttg cgcggggtc acatcgtcgg tgcctggata gcggggccagc tccggcagca
 18061 gcgcggcgag ctgcgggacc aaccattccg gggccgggtc atggcggagc ttgacggcga
 25 18121 agtcacgac gccggggcgc acatcctgat caccgtgta gcgcgccgcg gcaagcggcg
 18181 tagtgtctag actgccaca gcgtcaaca gtatggggcc ggtgtgcggg ccaagaatcc
 18241 agagcaccgc cgacgcgtg tctacgggc gacaaccgcg acatcacagc cagctaacag
 18301 ggcgtcggcg gtgatgatcg tcaggccaag cagctgtgcc tgggcgatga gcacacggtc
 18361 gaatgatgt cgtgatgat ccggaagctc tgccgtgcgc agtgtgtgc tggtaactg
 30 18421 acagcggcga cgtgccgag cggcgcatlc gatcgggac gtaagaagcc gatgctcgg
 18481 gcggcgggag cttgccgag cggtagttga tcgcgatc ccaggcactg gcggccgaca
 18541 agagaatgct gttgcggacg tctgaacaa tcgccgtgt ttggtgacg gcatccgag
 18601 ccaaacgttg gttcgtatga ggtagcgtt caccgtgaa agcgttcgag cagctcgtct
 18661 gacaacggag cgtccaatc gtcgggcagc cgggtacgc catggtcaat gcctaacgc
 35 18721 cgagctcat gaggatcag cggcacaagc ttgtaccg gtcgcggcg gcgggcaatc
 18781 tcaacctctg cccgccgtag acgagccgca gcagctcga caggcgtgtc ttgcctcgt

18841 gaacgccgac ccgcttcgca ggcgccgaga ctttcgcgtc gaccacctgc tcaccaaact
 18901 tcgcgatcat cgctgtatc cacagcgcca acgggtagcg gttgtccaa ccgttcgtc
 18961 aacgacaatg ggatcgtgac cgacacgacc gcgagcggga ccaatfgccc gcctctcca
 5 19021 cgcgcccccg cacggcgcgc atcgtcgccg ggtgaatcgc cgcagctggt gatctcgat
 19081 ctggacggca cgtcgaccga ctcgccgcgc ggaatcglat ccagcttcgc acacgcgctc
 19141 aaccacatcg gtgccccagt acccgaaggc gacctggcca ctacatcgt cggcccgccc
 19201 atcatgaga cgtcgcgcgc catggggctc ggcgaatccg ccgaggaggc gatcgtagcc
 19261 taccgggccg actacagcgc ccgcggttgg gcgatgaaca gcttgttcga cgggatacggg
 10 19321 ccgtcgttgg ccgacctgcg caccgocggt gtccggctgg ccgtcgccac ctccaaggca
 19381 gagccgaccg cacggcgaat cctcgccac ttcggaattg agcagcactt cgaggtcatt
 19441 cggcgccgca gcaccgatgg ctcgcgaggc agcaaggtcg acgtgctggc ccacgcgctc
 19501 gcgcagctgc ggcgcctacc cgaagggttg gtgatggtcg gcgaccgcag ccacgacgtc
 19561 gacggggccg ccgcgcacgg catcgacac gtggtggtcg gctggggcta cggcgccgcc
 15 19621 gactttatcg acaagacctc caccaccgtc gtgacgatg ccgccacgat tgacgagctg
 19681 agggaggcgc taggtgtctg atccgtgcga cgtcacatc gttgtacgg gcaacatctg
 19741 ccgttcgcca atggccgaga agatgttcgc ccaacagctt cgccaccttg gctctgggta
 19801 cgcgtgcgca gtgaccagtg cgggcaccgg gaactggcat gtaggcagtt gcgcgcgaga
 19861 gcggcgccgc ggggtgttgc gagccacgg ctacctacc gaccacgggg ccgcacaagt
 20 19921 cgcaccgcaa cactggcgg cagacctgtt gttggccttg gaccgcaac acgtcgcgt
 19981 gtgcgcgag ctgcgcgtcg aagccgccc ggtacgatg ctccggtcat tcgaccacg
 20041 ctgggaacc catgcgtcgt atgtcgagga tccctactat ggcgatcact ccgacttoga
 20101 gggagttctc gccgtcatcg aatccgcct gcccgccctg cagcactggg tcgacgaacg
 20161 tctcgcggg aacggaccga gttgatccc ccctagcgt tctgctcg gcccgctgg
 25 20221 ctggcgttgg cctcgtcgt gtgcgcgttc acctacctgt gctttacgt gtccgcgcg
 20281 tggcagctgg gaagaatgc caaaactca cgagagaacc agcagatcag gtattccctc
 20341 gacacccgc cggttccgt gaaaacctt ctaccacgc aggaatcgtc ggcgccggac
 20401 gcgcagttgg gccgggtgac ggcaaccgga cagtacctc cggacgtgca ggtgtcgcc
 20461 cgactgcgcg tgggtgaggg ggaccaggcg ttgaggtgt tggcccaat cgtgtgcag
 30 20521 ggcggacca ccgtctcgtt cgaccgtgga tacgtcggc ccgagttggg ctgcacgta
 20581 ccaccgatcc ccgcctgccc ggtgcagac gtgacctca ccgcggcgtt gctgactcc
 20641 gaaccgagcg tggcgggcaa agaccattc gtcagagac gcttcacga ggtgtatcg
 20701 atcaataccg gacaggtcgc cgcgtgcacc ggagtcacg tggctgggtc ctatctgcag
 20761 ttgatgaag accaaccggc cgggctcggc gtgctcgcg ttcgcactt agatccggg
 35 20821 ccgttctgt cctatggcat ccaatggatc tcgttcgga ttctggcacc gatcgcgtg
 20881 ggctatttgg cctacgccga gatccggcg gcgcgccggg aaaaagcggg gtgcgccaca

20941 ccggacaagc caatgacggt cgagcagaaa ctgcgtgacc gctacggccg ccggcggttaa
 21001 accaaatca cggccaatfc cgagcccc gcctggacca cccgcgacag caccacggcg
 21061 cggcgcatg cggccacctt gggcgaccgg ccgtcgccca aggtggggcg gatctgaac
 5 21121 tcattggtgt accgggtggg ccacccagc cgcacgtcaa gcgccccagc aaacgccgcc
 21181 tcgacgacg cgcggttggg gctgggatgg cgggcggcgt cgcgccgcca ggcccgtacc
 21241 gcaccgcggg gcgaccaccc gaccaccggc gcgcagatca ccaccagcac cgcctgcgcc
 21301 cgtgcgcca catagttggc ccagtcaccc aatcgtgctg cagcccaacc gaatcggaga
 21361 taacgcggcg agcggtagcc gatcatcgag tcagggtgt tgaatggcac atatccagc
 10 21421 accgcaggca cgcgcgtcga agccgccac agcagcggca ccacctggcg gtcggcggtg
 21481 ttctcgcca ccgactcag cgcggcacgc gtcaggcccg ggccgccacg ctggccgggg
 21541 tcacgcccg cagcgcagg cagcagccgt cgcgccgctt cgacatcgtc ggcgtccaac
 21601 aggtccgata tctggcgccc ggtgcgcgcc agcgaagttc cgccacgcg tcccagggtg
 21661 gccgtcggg tggccgccac ggccaggagc ctgccgggta gccgctgcag tggccggccg
 15 21721 agcaagccca ccgcgccgac cagcagcccg acgtgtaccg caccggcgac ccggccgtca
 21781 cggtaggtga tctgtccag ctggcgcc gcccgaccga acaggccac cggtagacct
 21841 cgtttgggtt cggcgaacac gacgtcagc aggcagccga tcagcacgc gacggccctg
 21901 gtctccagg tcgatgcaaa cactccggca gggtccgaca cgtgtcttac gctcagctac
 21961 ttatgacct ataccgcagc tatccacgt gaagcggcca gctaccggg tggccgacct
 20 22021 gtgaacccg gcggcaatgt tgttccggc agcgaatgt atcatcgag tggcagtgcc
 22081 ggggttcggg tatggcgtgc tggaaagccc ggtggacagc ggcaacgtct acaagcatcc
 22141 gttaagcgg gcccgacca ccggcaccta cctggcggtg gcgaccatcg ggcgggaatc
 22201 cgaccgagcg ctgaccggg gtgccttga cgtcgcgcac cggcaggttc ggtcgcagcg
 22261 ctgagccca gttcctata acgcttcga ccggaagttg cagctgtggg tggcgccgtg
 25 22321 tctgtaccg tacttctgg accagcacga gtttctgat ggccaccatg aagatgccac
 22381 cggcgacgcc gttaccaag acgccaacg gttagggacc acgtcgagg tggcgaggcg
 22441 gatgtgccc ccggaccggg tcgcttga cgaacttg aagcgtcgc ttgatggct
 22501 cgacatcgac gcggcgggtg gcgagcatct tcgggggtg gctcggtag cgttctcc
 22561 gttggcgttg cgcgcgttg ccgggcgtt caacctgtt ggcagcagcg gattctggc
 30 22621 accggagttc cgcgcgatga tgcagctga gttgtacag gccacgcagc gtcgttcga
 22681 gttgtactt tccgtctac ggtagccga ccgctgatt ccgcatcggt cctggaictt
 22741 cgtttaccag cttactgtt ggagacatgc gtttcgccc cgacacggcc gccgaatcgt
 22801 ctgatagagc ccggccgagt gtagcctga cagccgaca ccggcgcggt gtttcgctgc
 22861 gccaggttca cgtcggcga tctagagccg ccgaaacct acttctgggt tgcctccga
 35 22921 atcaacgtg tgaatgctc gagcagctca cgcatacgg cgcgcacgc atccaccgcg
 22981 gatacaggt cggccttgg cgcggcgagc tggtcgagc tcatgtggc caccggcggt

23041 gctgtctgtc ggcgccgctc gtcgcttga aaccacggc gtcacccac gaccacgaca
 23101 ctgccatata cggcgccccg ccgacaacga agcacagcta gccggtggcg cggcagggga
 23161 tcgaaccgcc gaccgtggtt ggtlaaaacc agagctctac cgctgagcta cgcgcccatg
 5 23221 accgccgcag gctacacgcc ttgcggccaa gcacccaaaa ccttaggcgg taagcgccgc
 23281 cagagcgctc gtcacacgcc gctgacgcg aactcacc ggctgtctca tctcgcgaa
 23341 ccgaatgata cctgaccgat cgaccacaaa ggtgccccg ttacgatgc cggcctgtc
 23401 gttgaagacg ccgtaggcct gactgaccgc gccgtgtggc cagaagtccg acaacagcgg
 23461 aaacgtgaat ccgtctctgc tcgccagat ctgtgtggtg ggtggcgggc ccacccaat
 10 23521 cgctagcgcg cgctgtctgt cgttctcaaa ctggcgagg tgatcagca actggtccag
 23581 ctgccctcgg cagatgcccg tgaacgcaa cggaaaagac accaacagca cgttcttgc
 23641 accccggtag ccgcgagggg tgacaagctg ctgattctg tcgcgaacg tgaagtcagg
 23701 ggcgggtgct ccgacgtica gcatcagcgc ttgcagccc gcgatttcgg ctgtaccaat
 23761 ctgtgtggcg tccagttgcc cagattgacc gacgagtgcg gcatcagccc agctgtggcg
 15 23821 gccgcctcgg caatctcggc gggcaatata tggccgggt gcccggtctt gggcgtcacc
 23881 acccaaatca caccgtctc ggcgagcggc ccgatcgc atcatcgggt gtccacaaa
 23941 tcgccgtcgc catcacgcca ccacaacagg acgacatcga tgacctgtc ggtgtcttca
 24001 tcgagcaact ctccccgca cgttcttgc atggccgcgc ggatgtgtc gtggtgtgtc
 24061 tcgtccacgc ccattctcgt gataagttg tctgttga tgcccaattt gcggcgtag
 20 24121 ttgaggcggt gatccgccgc gaccaccgtg gaacctctt cagtctcgc gggccatgtg
 24181 cacaccgtcg cgalggcat tatcgtcgca cagccagaac cgttccacc gccgcctca
 24241 gaaggcgcc acgcacattg tcaatgcctt tgtttgtg tctgtgagc gatcaaccg
 24301 ccggttgaat tccgtgtcg acgctgtcg accgatgca ttgccaccg cgcggcgccg
 24361 gtcgacatat gcgttgagcg catccccag ttgcgggac agcgcgcgcg tcagactgcc
 25 24421 tgagaccgtc gaggcactgt tgttgagcg gtcgatggc ggacctcgg tcggcccggt
 24481 gttgcgccc tgattgaacg cggccacgta ggcgttcaac ttgtgatgc cgtctctgt
 24541 ggtggccgcc agcgcgtaac acgaggtgc aatcgcttg gtcgtcagcg attgttggc
 24601 ctgcgaatcc cggatgctcg acgtgcgcc cgaaggcgac accgacggcg acaccagca
 24661 gcgttaggcc ggtgcgacgt tgggtcggc catggccgta ccgtcgttga cagtgttaca
 30 24721 tccgacgata ccatcagca gcagcgcgat gcagccgagc gccaggcgcg ctgcctggg
 24781 gagctcccc ccgtgcctgc gaggcacggc gcgccatccg atgagcacgg catgtgaggt
 24841 tactctgtcg cagcgcgacc gcgttggcg tgggtgtgic cgcatccgca gaaccgagcg
 24901 gagtgcggct atccgccgc gacgcccgtg cggcacgata gggggacgac catctaaaca
 24961 gcacgcaagc ggaagccgc cactacagg agtagtgcgt tgaccacga ttctgccgc
 35 25021 cacgatctgg cccaaaac aaacagcgca agcgaaccg accgattcg ggtgatccg
 25081 gagggtgtg cgtcgtattt gcccgacatt gatccgagg agacctcga gtgctgtgag

25141 tccttfgaca cgctgctgca acgctgcggc ccgtcgcggg ccgctacct gatgtgcgg
 25201 ctgctagagc gggccggcga gcagcgggtg gccatcccg caltgacgtc taccgacat
 25261 gtcaacacca tcccaccga gctggagccg tggttcccg gcgacgaaga cgtcgaacgt
 5 25321 cgttatcgag cgtggatcag atggaatcgc gccatcatgg tgcaccgtgc gcaacgaccg
 25381 ggtgtggcgg tgggtggcca tatctgacc tacgcgtcgt ccgcggcgt ctatgaggtc
 25441 gggttcaacc acttctccg gcgcaagtc caccggcgcg gcgcgcatca ggtgttcac
 25501 cagggccacg ctccccggg aatctacgc cgcgccttcc tcgaaggcg gttgaccgcc
 25561 gagcaactcg acggattcgc ccaggaaac agccatgtcg gcgcggggt ccgctccat
 10 25621 ccgcaccgc ggctcatgcc cgacttctgg gaattccca ccgtgtcgat ggtgtgggc
 25681 ccgctaacg ccatctacca ggcacggctc aaccactat tgcatgaccg cggatcaaa
 25741 gacacctccg atcaaacgt gtggtgttt ttggcgacg gcgagatgga cgaaccggag
 25801 agccgtgggc tggccacgt cggcgcgtg gaaggcttgg acaactgac ctctgtatc
 25861 aactgcaatc tgcagcgact cgacggcccg gtgcgcgca acggcaagat catccaggag
 15 25921 ctggagtcgt tcttcgcgg tgcggctgg aactgcatca agtgggtgtg gggcgcgaa
 25981 tgggatgcc tgcgcacgc cgaccgcgac ggtgcgctgg tgaattaat gaatacaaca
 26041 ccgfatggcg attaccagac ctataaggcc aacgacggcg gctacgtgcg tgaccactc
 26101 ttgcgcccg acccacgcac caagcgcgtg gtggagaaca tgagcgacca ggatatctgg
 26161 aacctcaaac gggcggcga cgtatccgc aaggtttacg ccgcctaccg cgcgcgcgtc
 20 26221 gaccacaagg gacagccgac ggtgatcgt gccaaacca tcaaggctga cgcgttggc
 26281 aagcatttcg aaggacgcaa tgcacccac cagatgaaaa aactgacct ggaagacct
 26341 aaggagtttc gtgacacga gcggattcgc gtcagcgacg ccagcttga agagatccg
 26401 tacttgccg cctactacca ccccgccctc aacgccccg agattcgtta catgctcgac
 26461 cggcgccggg ccctcggggg ctgttccc gagcgaggga ccaagtccaa agcgtcgacc
 25 26521 ctccgggtc gcgacatcta cgcgcgctg aaaaaggct ctggcgacca ggaggtggcc
 26581 accaccatcg cgacggtcg caggttcaa gaagtgttc gcgacaagca gatcgggccg
 26641 cggatagtc cgtatctcc cgacgaggcc cgcaccttcg ggtggaactc ctgttccc
 26701 tcgtataaga tctataaccg caatggccag ctgtataccg cggttgacgc cgactgatg
 26761 ctggctcata agggagcgca agtcgggcag atcctgcacg agggcatcaa cgaagccggg
 30 26821 tcgggtggct cgttcatcgc ggcggcacc tcgtatcgca cgcacaaga accgatgat
 26881 cccatttaca tcttactc gatgttcggc ttccagcgca ccggcgatg ctctgggccc
 26941 ccggccgacc agatggctcg aggttctgt ctcggggcca ccgccggcg caccacctg
 27001 accgctgagg gcctgcaaca cgcgcaggt cactcgttgc tgcgtgccg caccaaccg
 27061 gcggtggtg cctacgacc ggccttcgcc tacgaatcg cctacatgt ggaagcgga
 35 27121 ctggccagga tctcgggga gaaccggag aacatctct tctatcatc cgtctacaac
 27181 gagccgtacg tgcagccgcc ggaagccggag aattcgtac ccgaggcggt gctcgggggt

27241 atctaccgct atcacgcggc caccgagcaa cgcaccaaca aggcgcagat cctggcctcc
 27301 ggggtagcga tgcgccggcg gctcgggca gcacagatgc tggccgccga ttgggatgtc
 27361 gccgccgacg tgtggctcgt gaccagttgg ggcgagctaa accgcgacgg ggtggccatc
 5 27421 gagaccgaga agctccgccca ccccgatcgg ccggcggggc tgcctctagt gacgagagcg
 27481 ctggagaatg ctccggggccc ggtgatcgcg gtgtcggact ggaatgcgcg ggtccccgag
 27541 cagatccgac cgtgggtgcc gggcacatac ctacgttgg gcaccgacgg gticggctti
 27601 tccgacactc ggcccgcgcg tcgccgtac ttaacaccg acgccgaatc caagttggtc
 27661 gcggttttgg aggcgttggc gggcgacggc gagatcgacc catcgttggc ggtcgcggcc
 10 27721 gccgccagtt accggatcga cgacgtggcg gctgcgcccg agcagaccac ggaatccggt
 27781 cccggggcct aacgccggcg agccgaccgc ctttggccga atcttcaga aatctggct
 27841 agcttttagg agtgaacgac aatcagttgg ctccagtgc ccgccgagg tcgccgtcg
 27901 aactcgtgga cactgtgcc gattcgtgc tgcggcggti gaagcagtac tcgggccggc
 27961 tggccaccga ggcagtttgc gccatgcaag aacggttgc gtcttcgcc gacctagaag
 15 28021 cgtcccagcg cgcagcgctg gcgctgtgtg tgcagacggc cgtcttcaac tctctgaat
 28081 ggaatgcaga ccgcacagt gacgtcggtc ataccgcga ggaattcgag ctgtgtcccc
 28141 aggatctgac gcgacggatc gcgctgcgcc agaccgtgga catgtgcggc gtcacatgg
 28201 agttcttcca agaagtcgtg cccctgctc ccgcttccga agagcagttg accgccctca
 28261 cggtgggcat ttgaataac agccgcgacc tggcatfacc cgccgccagc gcctacgccg
 20 28321 atcgcggcca ggcacgaggc acctgggaca gccgatgga ggcacgctg gtgagacggg
 28381 tggtagcgcg cgacaccggt cccgagctgc tgtcccggcg ggccgcgctg aattgggaca
 28441 ccaccgcgcc ggcgaccgta ctgttgggaa ctccgcgcgc cgttccaaat ggtctcaaca
 28501 gcgacggcga cagcgagcgg gccagccagg atgtccgcga caccgcggct cggccagggc
 28561 gcgctgcgct gaccgacgtg caccgcacct ggtctgttgc gatcgtctcc ggccagctgt
 25 28621 gcgaaccga gaagtcttc aaagacctgc tggcagcatt cggcagccc ccggtgttca
 28681 tcggcccccac ggcgcccatg ctgaccgcgg cgaccgcag cgtagcgag gcgactccc
 28741 ggaatgaacg cgtcgccggc tggcgcgagg cgccgcggcc cgtgctggt agggaaactt
 28801 tgcccgaaag cgccctgatg ggcgacgcct cggcgatcgt ggccctgcat accgacgtga
 28861 tgcggccctc agccgatgoc ggaccgacgc tcatcgagac gctagacga tatctggatt
 30 28921 gtggcgcgcg gattgaagct tgtccagaa agttgttct tcatccaaac acagtgcggt
 28981 accggctcaa gcgatcacc gacttcaccg ggcgcgatcc caccagcca cgcgatgcct
 29041 atgtccttgg ggtggcgggc accgtgggtc aactcaacta tccgagccg cactgaagca
 29101 tcgacagcaa tgccgtgtca tagattccct gcccggtcag aggggggtcca gcaggggccc
 29161 cggaaagata ccaggggcgc cgtcggacgg aaagtgatcc agacaacagg tcgggggacg
 35 29221 atctcaaaaa catagcttac agggccgttt tgttggttat atacaaaaa ctaagcagag
 29281 gttcataatc tgttacaccg cgcaaacccg tcttcacagt gtctcttag acacgtgatt

29341 gcgttgctcg caccggaca ggggtcgcaa accgaggaa tgtgtgcgc gtggttcag
 29401 ctgcccggcg cagcggacca gatcgggcg tggtcgaaag ccgtgatct agatcttgc
 29461 cggctgggca ccaccgctc gaccgaggag atcaccgaca ccgcggtcgc ccagccattg
 5 29521 atcgtgcgcg cgactctgt gggccaccag gaactggcgc gccgatgcgt gctgcggcgc
 29581 aaggacgtca tctgtgccgc cactccgtc ggcgaaatcg cggctacgc aatgcgggt
 29641 gtgatagccg ccgacgacgc cgtcgcgtg gccgccacc gccgcgcga gatgccaaag
 29701 gcgtgcgcca ccgagccgac cggcatgtct gcgtgtcgc gcggcgacga gaccgagtg
 29761 ctgagtcgcc tcgagcagct cgacttgct ccggcaaac gcaaccgcgc cggccagatc
 10 29821 gtgcgtccg gccggctgac cgcgttgag aagtcgccg aagaccgcc ggccaaggcg
 29881 cgggtgcgt cactgggtgt gcgggagcg ttccacacc agttcattgc gcccgactt
 29941 gacgctttg cggcgccgc ggccaacat gcaaccgcc accccaccgc cagctgctg
 30001 tccaaccgcg accgggaagc ggtgatcc gcggccggcg cgatggacac cctgtctcc
 30061 cagctcacc aaccgtgcg atgggacctg tgcaccgca cgtcgcgca acacacagtc
 15 30121 accgcatcg tggagtccc cccgcgggc acgttagcg gtatcgcaa acgcaact
 30181 cggggggttc cggcacgcgc cgtcaagta ccgcagacc tggacgagct ggcaaaccta
 30241 taaccgcgga ctggccaga acaaccacat acccgtcagt tgcatttga cacaacatat
 30301 tacgaaggga agcatgctgt gcctgtcact caggaagaaa tcattgccg tatcgccgag
 30361 atcatgaaag aggtaacgg tatcgagccg tcgagatca ccccgagaa gtcttctgc
 20 30421 gacgacctgg acatcgact cgtctgatg gtccgatcg ccgtgcagac cgaggacaag
 30481 tacggcgta agatccccga cgggacctc gccglttgc gtaccgtgc tgacgttgc
 30541 gcctacatcc agaagtcga ggaagaaac ccggaggcgc ctacggcgtt gcgcgcgaag
 30601 attgagtcgg agaaccgcc tgcgttgcc aacgticagg cagggttga ggccgagtc
 30661 aagttagtca gccctccacc gctaattgcg gtttccag cgttgtgtg accgcgtca
 25 30721 cagcgacgac gtgatctcg ccggacatcg agagcacgt gaagggtgt ttggccggcg
 30781 agagcggcat ccacgcact gaagacgagt tctcaccaa gtggatcta gcgtcaaga
 30841 tcgggggtca cctcaaggat ccgtcgaca gccacatgg ccgactcgac atcgacgca
 30901 tctgtactgt ccagcggatg ggcaagttgc tggcgggaca gctatggag tcgccggca
 30961 gcccgagggt gatccagac cgttcgccg ttgtgtcgg caccgtcta gttggagccg
 30 31021 agaggaattg cgagagctac gacctgatg atcggcgcg ccccggaag gtttcccgcc
 31081 tggccgtica gatgatcat ccaacggtg ccgcggcgt gatgggtcgc cagcttgggg
 31141 ccgcggccgg ggtgatgacc ccgtgtcgg cctgtctgc gggctcgaa gcgatcgcc
 31201 acgcgtggcg tcagatctg atgggcgac ccgacgtgc cgtctgcgc ggtgtcgaag
 31261 gaccatcga ggctgtgcc atcggcggt tctcatgat gcgggcatg tcgaccgca
 31321 acgacgagcc tgagcgggccc tccggccgt tcgacaagga ccgcagcgc ttgtgttcg
 35 31381 gcgagccggc tgcgtgatg ctcatcgaga cggaggagca ccgcaagcc cgtgcgcga

31441 agccgttgcc ccgattgctg ggtgcccgtg tcacctcgga cgccttcat atgttgccgc
 31501 ccgcggccga tgggtttcgt gccggtaggc cgaatgactg ctgcctggag ctggccgggt
 31561 tgtcgccgcg ggacatcgac cactgaacg cgcacggcac ggcgacgcct atcgccgacg
 5 31621 ccgcggaggc caacgccatc cgcgtccgcg gttgtgatca ggccgcggtg tacgcgccga
 31681 agtctgctcg gggccactcg atcggcggcg tgggtgcgtc cgaatcggtg ctacgggtgc
 31741 tgacgtcgcg cgacggcgtc atcccgccga cctgaacta cgagacacc galcccgaga
 31801 tcgacctga cgtcgtgcgc ggcaaccgc gctatggcga ttaccgtac cgactcaaca
 31861 actcgttcgg gttcgccgc cacaatgtg cgttgcctt cggcggttac tgaagcacga
 10 31921 catcgccggt ccgagggccc gagggtgggg tccccgcct tgcggggcg agtcggaccg
 31981 atatggaagg aacgttcgca agaccaatga cggagctggt taccgggaaa gcctttccct
 32041 acgtagtctg caccggcatc gccatgacga ccgcgtcgc gaccgacgcg gagactacgt
 32101 ggaagtgtt gctggaccgc caaaggcgga tccgtactct cgaatgacca ttgctcgagg
 32161 agttcgacct gccagtgcg atcggcggac atctgttga ggaattcgac caccagctga
 15 32221 ccgcgatcga actgcgcggc atgggatacc tgcagcggat gtccaccgtg ctgagccggc
 32281 gccttgggga aatgccggc tcaccggagc tggacaccaa tcatgtatg gttctcatcg
 32341 gcaccggcct gggttcgcc gaggaactgg tcttcagtta cgacgatag ccgcgtcgcg
 32401 gaataaggc ggtctcgcc ctagaccgtc agaagtacat gcccaacggc gccgcgccgg
 32461 cggctgggtt ggaacggcac gccaaaggcc ggggtatgac gccggtatcg cgtgtcgcat
 20 32521 ccggcgccga ggccatcgcc cgtcgtgac agcagatgt cttggggag agccgatccg
 32581 ccatctcgcg cggcgtggag accagatcg aagcgtgccc catcgccggg ttgctcaga
 32641 tgcgcatcgt gatgtcacc aacaacgacg acccgcggc tgcgtgccgc ccatcgaca
 32701 gggaccgcga cggctttgtg ttccggcagg cggcgcgcct tctgtgac gagaccgagg
 32761 agcacgccaa ggcacgtggc gccaatcc tggccggat catggcgcc agcatcacct
 25 32821 ccgatggctt ccacatgtg gccccggacc ccaacgggga acgcgcggg catcgogatta
 32881 cgcggcgcat tcagctggcg ggcctcgccc ccggcgacat cgaccacgtc aatgcgcacg
 32941 ccaccggcac ccaggtcgcc gacctggccg aaggcaggc catcaacaac gccttgggcg
 33001 gcaaccgacc gggcgtgtac gccccaaagt ctgccctcg ccatcgtgtg ggcgcggctg
 33061 gcgcggtcga atgatcttg acgtgctcgc cgttcgcga tcaagtgtac ccgcgcacac
 30 33121 tgaatctggt aaacctgat ccgagatcg atttgacgt ggtggcggtt gaaccgcgac
 33181 cgggcaatta ccggtatgag atcaataact cgttcggatt cggcgccac aacgtggcaa
 33241 tgccttcgg acggtactaa acccagcgt tacgcgacg gagacctcg algacaata
 33301 tggcccccga ggcggttgcc gagtgcctc accccgcga tccgtgtg cggctgagca
 33361 acttcttcga cgaaggcagc gtggaaatgc tgcacgagcg tgaccgtcc ggagtgtcgg
 33421 ccgcggcggg caccgtcaac ggtgtgcgca ccatcgctt ctgacaccg ggcaccgtga
 35 33481 tggcgcgcc catggcgctg gagggtgca cgcacatct caacgcctac gacactgcca

33541 tcgaagacca gaggcccatc gtgggcatct ggcatcggg ttgtgcccg ctggctgaag
 33601 gtgtcgggc gctgcacgcg gtaggccagg tgttcgaagc catgatccg cggtccggt
 33661 acatcccga gatctgggtg gtctcggtt tcgcccgccg cggcgccgc tacggaccgg
 5 33721 cggtgaccga cgtcgtctc atggcgccgg aaagccgggt gtctgcacc gggcccgacg
 33781 ttgttcgcag cgtcaccggc gaggacgtcg acatggctc gtctcgtggc ccggagacc
 33841 accacaagaa gtccgggggtg tggccatcg tcgccacga cgaactgat gctctagacc
 33901 gtggcgcccg gttgtcggg ttgtctgcc agcaggggca ttctgatcg agcaaggcc
 33961 aggcgggtga caccgacatc cagcgctcg tccgggaatc ctccgacgt gctctagacg
 10 34021 tgcgtccgat cgtgacggcg atctctgatg cggacacacc gtccgacgag ttccaggcca
 34081 attggcgcc gtcgatggtg gtccggctgg gtccgctgtc ggtcgcacg gtgggtgtac
 34141 tggccaacaa cccgtacgc ctggcgcggt gctgaactc cgaagcgca gagaaggcag
 34201 cgcgttctgt cggcgtgtgc gacgcgttc ggattccgt ggtgtgtgtg gtcatgtgc
 34261 cgggtctatc gcccggtgtc gaccaggagt ggggtggcgt ggtgcgcgt ggcgccaagt
 15 34321 tgctgcacgc gttcgcgag tgcaccgttc cggcggtcac gctggtcac cgaagaacct
 34381 acggcggggc atacattgcg atgaactccc ggtcgttgaa cgcgaccaag gtgttcgct
 34441 ggccggacgc cgaggctcgc gtgatggcg ctaaggcggc gtcggcctc ctgcacaaaga
 34501 agaagtggc cgccgctccg gacacgaac gcgaagcgt gcacgaccag ttggcgccg
 34561 agcatgacgc catcgccgc ggggtcgaca gtgcgttga catcggtgtg gtgcacgaga
 20 34621 agatcgacc ggcgcatact cgcagcaagc tcaccgagc gctgcgcag gctccggcac
 34681 ggcgcggccg ccacaagaac atcccgctgt agttctgacc gcgacgacg gcagaatgc
 34741 acgcgcgagg tcgcgcgtg gcgaltctg gtctgtctgc cagttatcc cagcggtggc
 34801 ttgtcaacgc gaggcgctcc tcgcatgtc ggacggtgcc taccgacgc ctaacaattc
 34861 tcgagaaggc cggcggttc gccaccacc cgcaattgt caccgtcatg acccgccaac
 25 34921 agtctgacgt ccaagtga aaacggcgcc tctgtcgtg ttgttatcg gtctacggc
 34981 cacaagagcc ggacgtgtg ggccgcttgg cggctctga tctgtcatg gggggcgacg
 35041 ccgtcgcgtg tctgggcacc gccgcgcgt tgtatgatt cgacacggaa aacaccgtc
 35101 ctatcatat gctcatgcc ggagtaagga tggcgccac ggtcgtctg atgtgccac
 35161 aacgcgtcgg tgcccgctc caacgggtgt caggtcgtc cgcgaccgc ccgcgatga
 30 35221 ctccgttga ggtcgcacga cagttgcgc gccgcgggc gctggccac ctgcagccg
 35281 cactacggtc aatgcgtcgc gctgcagtg aaattgaaaa cgccgttgtg gagcagcgag
 35341 gcccgcgagg catgctcgc gcgcgcgaac tcttaccct cgcgcagga cgcgcgggaat
 35401 cggccatgga gacgagcgt cggtcgtca tcatgacca cgggctccg ttgccgaac
 35461 ttcaataccc gatacacgc caggtgtgtg aaatgtggc agtcgactc gctcgcccg
 35 35521 acatcgctc cgcggccgaa tacgaaagca tcgagtggca cgcggggacc gcggagatg
 35581 tgcgcgcaa gacacgctg gccaaagctc aagagctcgg gtggacgatt gtccgattg

35641 tcgtcgacga tctcagacgc gaacccggcc gcctggcggc cgcctcgcgc cgcacacctcg
 35701 accgcgcgcg tatggccgcg tgaccgctgg tgagcagacg cagagtcgca ctggcgccgg
 35761 cgcagtcgca ctctgcctct gctcgcctc aacgcgtgag gaactccita gccacggcga
 5 35821 ctacgcgcct gcgatccgt ggaccagac cgaaccgggt ccggcgctcg aggatctgt
 35881 ccacatccag cgcgccctca tgggtcacgc cgtattcgaa ctccgccgcg gtcacgtoga
 35941 tgccgtcggc gaccggctcg gtggcgccgt cacatgtggc ggcgcgacgc acgttggccg
 36001 cctcgccccc gtaccgcgc accagcgact cgggcaatcc ggcgcccgat ccggggcgccg
 36061 gccacgggtt cgcgggtgcg ccgatcagcg gcaggttgcg agtcgcgcac ttgcggctc
 10 36121 gcagggttcg cagcgtgatg gcgcgatca gcacalcctc tgccatgtag cgttatccg
 36181 tcagcttgc gccgaccaca ctgatcacgc ccgacggcga tcaaaaaa cgtgtgtac
 36241 gcgaacgctc ggcggtcgg ccttgacac cagcaccgcc ggtgtcgait agcgcccgca
 36301 atccgcata ggacccgatg acatccttgg tgcgaccgc cgtcccaat gcgggttca
 36361 ccgatccag caggaaactg atctcttcg aagacggttgc tggcacatcg ggaatgggc
 15 36421 cgggtgcgtc ttctcgtc agcccgagat agatccggcc cagctcctcg ggcatggca
 36481 acacgaagcg gttcagctca ccggggatcg gaatggtcag cgcggcagtc ggaatggca
 36541 acgacttcgc gtcgaagacc agatgtgtgc cgcgcctggg gcgtagcctc agggcagcgt
 36601 cgaatcacc ccccccacgc ccgcgcgtg tgaatgacgc acgcgccgac agcgcgacag
 36661 actgcgggtg gcgccggtcg gtcaatcca ccgaagtgc ggtgacatc gacgcgccca
 20 36721 cgtaagttag gatcgggcg ccgtgctggg ccgcggtgcg cgcgacggcc atgaccaggc
 36781 gggcgtctc gatcaattgc ccgtctgac cgagcagacc accgtcgagg ccgtccgcc
 36841 gaacgggtgg agcaatctc accaccgtg acgcgggat tcggcgcat cggggcaacg
 36901 tcgcccccg cgtaccgct agcaccgca aagcgtcgc ggccaggaaa ccggcacgca
 36961 ccaacgccc cttgtgtga ccatcgacg gcaacaacgg gaccagttgc ggcattggat
 25 37021 gcacgagatg aggagcgttg cgtgtcatca ggaatccgc ttgcagcgcg ctgcgccgg
 37081 cgaatccac gttccgctg gccagatagc gcagaccgcc gtgcaccaac ttgagctcc
 37141 agcggttgtt gcgaacgcc agatcatgt ttccaccaaa ggccaccgtc agaccggg
 37201 tggcagatc taaggcaatg ccaacaccgg taatgccgcc gccatcacg atgacgtga
 37261 gtgcgccacc gtcggccagt gcgtcaggt cggcgagcgc acgcgccgcg ttgagtgcg
 30 37321 ccgagtgagg catcacaca aatatccgtt cagtcgtgg gtaagtctg tggccagcgc
 37381 gggcgcaatc aggatcgaat cgacgatgc cgggactgg atggtcgact gggcgatcag
 37441 caacacatg gtcgccagtc gacgagcgtc gccggagcgc acactgcgc accgtgcgc
 37501 cactgtcagc cggcgggcca accctcgat caggacgtgc tggctgggtc cgaggcgtc
 37561 ggtgatgtac accctggcca gtcgcagtg catgaccgac atgatcagat cgtaccccg
 35 37621 caaccggtc gccaccgca caatctgct taccacgct tccggtctg cccgtcag
 37681 gggcacctc cgcagcagc cggcgatatg gctggtcagc atggacgcca tgacgaccg

37741 gggtgccggc cagcgacggt atacggtcgg gcggctcacg cccgcgcgcc gggcgatc
 37801 ggcaagtgtc acccggcca cgcgtaatc gacgacgag ctgcgcgtg ccgcgagat
 37861 acgaccaccg gtatccgcgc ggtaactt cattgacagc atgtgtaata ctgtaacgcg
 5 37921 tgactcaccg cgaggaaact cttcaccca tgaatggga cgcgtggga gatccgcgc
 37981 cggccaagcc actttctgat ggcgtccggt cgtgtctgaa gcaggtgtg gccttagcgg
 38041 actcgggagca gcccaactc gaccccgccg aggtgcagct gcgccgtcc gcctgtcgg
 38101 gggcgagaca (SEQ ID NO: 24)

6.9. X-linked Inhibitor of Apoptosis Protein ("XIAP")

10 GenBank Accession # U45880:
 1 gaaaaggtag acaagtccta ttccaagag aagatgactt ttaacagtt tgaagatct
 61 aaaacttgtg tacttcgaga catcaataag gaagaagaat ttgtagaaga gttaataaga
 121 ttaaaactt ttgctaatt tccaagtggt agtctgttt cagcalcaac actgacgaca
 15 181 gcagggtttc ttatactgg tgaaggagat accgtgcggt gctttagtgt tcatgcagct
 241 gtatagatag ggcaataggg agactcaga gtgggaagac acaggaaaagt alccccaat
 301 tgcagattta tcaacggctt ttacttgaa aatagtgcca cgcagctac aaattctggt
 361 atccagaatg gtcagtacaa agttgaaac tatctggaa gcagagatca tttgctcta
 421 gacagccat ctgagacaca tgcagactat cttttgagaa ctgggcaggt tgtatataa
 481 tcagacacca tatacccgag gaacctgcc atgtattgtg aagaagctag attaaagtc
 20 541 ttcagaaat ggccagacta tgcacacta accccaagag agttagcaag tgcctgactc
 601 tactacacag gtattgtga ccaagtgcag tgccttgtt gtgttggaac actgaaaaat
 661 tgggaacctt gtgatctgc ctgttcagaa cacaggcgac actttctaa ttgctcttt
 721 gttttgggc ggaatctaa tatcgaagt gaatctgat ctgtgagtic tgalagaaat
 25 781 ttcccaaat caacaaatc tccaagaaat ccatcaatgg cagattatga agcacggatc
 841 ttacttttg ggacatggat atactagtt aacaaggagc agcttgcaag agctggattt
 901 tatgctttg gtgaaggiga taaagtaag tgccttcatg gtggaggagg gctaactgat
 961 tggagccca gtgaagacc ttgggaacaa catgctaact ggtatccagg gtgcaaatat
 1021 ctgttagaac agaagggaca agaataata acaatattc atttaacta ttactgtgag
 1081 gagtgtctgg taagaactac tgagaaaaca ccatcactaa ctagaagaat tgaatgatac
 30 1141 atcttccaaa atcctatggt acaagaagct atacgaatgg gggtcagtt caaggacatt
 1201 aagaaaataa tggagggaaa aattcagata tctgggagca actataaatc acttgaggtt
 1261 ctggttcgag atctatgaa tgcctagaaa gacagtatgc aagatgagtc aagtcagact
 1321 tcaitacaga aagagattag tactgaagag cagctaaggc gcctgcaaga ggagaagatt
 1381 tgcaaaatct gtatggatag aaatafctct atcgttttg ttctgtgg acatctagtc
 35 1441 actgttaaac aatgtgctga agcagttgac aaggtgccca tgtgtcacac agtctaact

1501 ttcaagcaaa aaatitttat gicttaact aactctatag taggcattgt atgtgtict
 1561 tattaccctg attgaatgt tgaatgaac tgactttaag taatcaggat tgaattccat
 1621 tagcattgc taccaagtag gaaaaaaaat gtacatggca gtgtlltagt tggcaatata
 5 1681 atctttgaat tcttgattt ttcagggtat tagctgtatt atccatttt ttactgttta
 1741 ttttaattgaa accatagact aagaataaga agcatcatac tataactgaa cacaaatgtg
 1801 attcatagta tactgattta attictaagt gtaagtgaat taatcatctg gattttttat
 1861 tcttttcaga taggcttaac aaatggagct tctgtatat aaatgggag attagagtta
 1921 atctcccaa tcacataatt tgttttgtgt gaaaaggaa taaatgttc catgctgtgt
 10 1981 gaaagataga gattgtttt agaggttgtt tgtgtgttt taggattctg tccatttct
 2041 tgtaaggga taaacacgga cgtgtgcgaa atatgttgt aaagtattt gccattgtg
 2101 aaagcgtatt taatgataga atactatcga gccaacatgt actgacatgg aaagatgca
 2161 gagatatgt aagttaaaa tgcaatggc gggacactat gtatagtctg agccagatca
 2221 aagtatgat gtgttaata tgcataaac gagagattg gaaagatata cacaaactg
 15 2281 taaatgttg ttctcttcg gggagggggg gattggggga gggggcccg aggggttta
 2341 gaggggcctt ttaacttcg actttttca ttittcttg ttcgatttt ttataagtat
 2401 gtgaccccg aagggtttta tgggaactaa catcagtaac ctaaccccg tgactalcct
 2461 gtgctcttc tagggagctg tttgtttcc caccaccac ccttcctct gaacaaatgc
 2521 ctgagtctg gggcacttg (SEQ ID NO: 25)

20

General Target Region:

Internal Ribosome Entry Site (IRES) in 5' untranslated region:

5'AGCUCCUAUAACAAAAGUCUGUUGCUUGUUGUUUCACAUUUUGGAUU
 UCCUAAUAUAAUGUUCUCUUUUUAGAAAAGGUGGACAAGUCCUAUUU
 25 UCAAGAGAAG3' (SEQ ID NO: 26)

25

Initial Specific Target Motif:

RNP core binding site within XIAP IRES

5'GGAUUUCCUAAUAUAAUGUUCUCUUUUU3' (SEQ ID NO: 27)

30

6.10. Survivin

GenBank Accession # NM_001168:

1 ccgccagatt tgaatcgcgg gaccctgttg cagaggtggc ggcggcggca tgggtgcc
 61 gacgttgcct cctgcctggc agccctttct caaggaccac cgcactctca cattcaagaa
 121 ctgcccttct ttggagggtc gcgcctgcac cccggagcgg atggccgagg ctgcttcat
 181 ccaactcccc actgagaacg agccagactt ggcccagtgt ttctctgtct tcaaggagct

35

241 ggaaggctgg gagccagatg acgaccccat agaggaacat aaaaagcait cgiccgggtg
 301 cgctttcctt tctgtcaaga agcagtttga agaattaac cttggtgaat ttttgaact
 361 ggacagagaa agagccaaga acaaaattgc aaaggaacc aacaataaga agaaagaatt
 421 tgaggaaact gcgaagaaag tgcgccgtgc catcgagcag ctggctgccca tggattgagg
 481 cctctggccg gagctgcctg gtcccagagt ggctgcacca ctccagggg ttatccctg
 541 gtgccaccag ccttctgtg ggccccttag caatgtcta ggaaaggaga tcaacattt
 601 caaattagat gtttcaactg tctcctgtt ttgtctgaa agtggcacca gaggtgcttc
 661 tgcctgtgca gcgggtgctg ctggtaacag tggctgttc tctctctc tctctttt
 721 gggggctcat ttttgcgtt ttgattccc ggcttaccag gtgagaagt agggaggga
 781 aaggcagtg ccttttgc agagctgaca gctttgtcg cgtgggcaga gccttcaca
 841 gtgaatgtg ctggacctca tgttgttgag gctgtcacag tctgagtg ggacttgga
 901 ggtgcctgtt gaatctgagc tgcaggttcc ttatctga cacttgccc tctcagagg
 961 acagttttt tgttgttg ttittttt ttittttt ggtagatga tgcattgtg
 1021 gtgatgagag aatggagaca gactccctg ctccttact gtttaacaac atggctttct
 1081 tattttgtt gaattgttaa ttacagaat agcacaaact acaattaaa ctaagcaca
 1141 agccattota agtcatggg gaaacgggt gaactcagg tggatgagga gacagaatag
 1201 agtgatagga agcgtctgac agatactct ttgccactg ctgtgtgatt agacagccc
 1261 agtgagccgc ggggcacatg ctggccgctc ctccctcaga aaaaggcagt ggcctaaatc
 1321 ctttttaaat gacttgctc gatgcttgg gggactggct gggctgtgct aggcctgtg
 1381 tctgtcagcc caacctcac atctgtcac ttctccacac gggggagaga cgcagtcgc
 1441 ccaggtcccc gctttcttg gaggcagcag ctccgcagg gctgaagtc ggcgtlaagt
 1501 gatggatttg atcgcctc ctccctgca tagagtgca ggtgggattg ttacagctc
 1561 gctggaaac tctggagtc atctcgctg ttctgagaa ataaaaagcc tgtcatttc (SEQ ID NO: 28)

25

7. EXAMPLE: IDENTIFICATION OF A DYE-LABELED TARGET RNA BOUND TO SMALL MOLECULAR WEIGHT COMPOUNDS

The results presented in this Example indicate that gel mobility shift assays can be used to detect the binding of small molecules, such as the Tat peptide and gentamicin, to their respective target RNAs.

7.1. Materials and Methods

7.1.1. Buffers

Tris-potassium chloride (TK) buffer is composed of 50 mM Tris-HCl pH 7.4, 20mM KCl, 0.1%Triton X-100, and 0.5mM MgCl₂. Tris-borate-EDTA (TBE) buffer is

composed of 45 mM Tris-borate pH 8.0, and 1 mM EDTA. Tris-Potassium chloride-magnesium (TKM) buffer is composed of 50 mM Tris-HCl pH 7.4, 20mM KCl, 0.1% Triton X-100 and 5mM $MgCl_2$.

5

7.1.1. Gel retardation analysis

RNA oligonucleotides were purchased from Dharmacon, Inc, Lafayette, CO). 500 pmole of either a 5' fluorescein labeled oligonucleotide corresponding to the 16S rRNA A site (5'-GGCGUCACACCUUCGGGUGAAGUCGCC-3' (SEQ ID NO: 29);
10 Moazed & Noller, 1987, Nature 327:389-394; Woodcock *et al.*, 1991, EMBO J. 10:3099-3103; Yoshizawa *et al.*, 1998, EMBO J. 17:6437-6448) or a 5' fluorescein labeled oligonucleotide corresponding to the HIV-1 TAR element TAR RNA (5'-GGCGUCACACCUUCGGGUGAAGUCGCC-3' (SEQ ID NO: 30); Huq *et al.*, 1999, Nucleic Acids Research. 27:1084-1093; Hwang *et al.*, 1999, Proc. Natl. Acad. Sci. USA 96:12997-13002) was 3' labeled with 5'- ^{32}P cytidine 3', 5'-bis(phosphate) (NEN) and T4 RNA ligase (NEBiolabs) in 10% DMSO as per manufacturer's instructions. The labeled oligonucleotides were purified using G-25 Sephadex columns (Boehringer Mannheim).
15 For Tat-TAR gel retardation reactions the method of Huq *et al.* (Nucleic Acids Research, 1999, 27:1084-1093) was utilized with TK buffer containing 0.5mM $MgCl_2$ and a 12-mer Tat peptide (YGRKKRRQRRRP (SEQ ID NO: 31); single letter amino acid code). For 16S rRNA-gentamicin reactions, the method of Huq *et al.* was used with TKM buffer. In 20 μ l reaction volumes 50 pmoles of ^{32}P cytidine-labeled oligonucleotide and either gentamicin sulfate (Sigma) or the short Tat peptide (Tat₄₇₋₅₈) in TK or TKM buffer were heated at 90 °C for 2 minutes and allow to cool to room temperature (approximately 24 °C)
20 over 2 hours. Then 10 μ l of 30% glycerol was added to each reaction tube and the entire sample was loaded onto a TBE non-denaturing polyacrylamide gel and electrophoresed at 1200-1600 volt-hours at 4 °C. The gel was exposed to an intensifying screen and radioactivity was quantitated using a Typhoon phosphorimager (Molecular Dynamics).

25

30

7.2. Background

One method used to demonstrate small molecule interactions with natural occurring RNA structures such as ribosomes is by a method called chemical footprinting or toe printing (Moazed & Noller, 1987, Nature 327:389-394; Woodcock *et al.*, 1991, EMBO J. 10:3099-3103; Yoshizawa *et al.*, 1998, EMBO J. 17:6437-6448). Here the use of gel mobility shift assays to monitor RNA-small molecule interactions are described. This
35 approach allows for rapid visualization of small molecule-RNA interactions based on the

difference between mobility of RNA alone versus RNA in a complex with a small molecule. To validate this approach, an RNA oligonucleotide corresponding to the well-characterized gentamicin binding site on the 16S rRNA (Moazed & Noller, 1987, Nature 327:389-394) and the equally well-characterized HIV-1 TAT protein binding site on the HIV-1 TAR element (Huq *et al.*, 1999, Nucleic Acids Res. 27: 1084-1093) were chosen. The purpose of these experiments is to lay the groundwork for the use of chromatographic techniques in a high throughput fashion, such as microcapillary electrophoresis, for drug discovery.

7.3. Results

A gel retardation assay was performed using the Tat₄₇₋₅₈ peptide and the TAR RNA oligonucleotide. As shown in Figure 1, in the presence of the Tat peptide, a clear shift is visible when the products are separated on a 12% non-denaturing polyacrylamide gel. In the reaction that lacks peptide, only the free RNA is visible. These observations confirm previous reports made using other Tat peptides (Hamy *et al.*, 1997, Proc. Natl. Acad. Sci. USA 94:3548-3553; Huq *et al.*, 1999, Nucleic Acids Res. 27: 1084-1093).

Based on the results of Figure 1, it was hypothesized that RNA interactions with small organic molecules could also be visualized using this method. As shown in Figure 2, the addition of varying concentrations of gentamicin to an RNA oligonucleotide corresponding to the 16S rRNA A site produces a mobility shift. These results demonstrate that the binding of the small molecule gentamicin to an RNA oligonucleotide having a defined structure in solution can be monitored using this approach. In addition, as shown in Figure 2, a concentration as low as 10ng/ml gentamicin produces the mobility shift.

To determine whether lower concentrations of gentamicin would be sufficient to produce a gel shift, similar experiment was performed, as shown in Figure 2, except that the concentrations of gentamicin ranged from 100 ng/ml to 10 pg/ml. As shown in Figure 3, gel mobility shifts are produced when the gentamicin concentration is as low as 10 pg/ml. Further, the results shown in Figure 3 demonstrate that the shift is specific to the 16S rRNA oligonucleotide as the use of an unrelated oligonucleotide, corresponding to the HIV TAR RNA element, does not result in a gel mobility shift when incubated with 10 µg/ml gentamicin. In addition, if a concentration as low as 10 pg/ml gentamicin produces a gel mobility shift then it should be possible to detect changes to RNA structural motifs when small amounts of compound from a library of diverse compounds is screened in this fashion.

Further analysis of the gentamicin-RNA interaction indicates that the interaction is Mg- and temperature dependent. As shown in Figure 4, when $MgCl_2$ is not present (TK buffer), 1mg/ml of gentamicin must be added to the reaction to produce a gel shift.

Similarly, the temperature of the reaction when gentamicin is added is also important. When gentamicin is present in the reaction during the entire denaturation/renaturation cycle, that is, when gentamicin is added at 90°C or 85°C, a gel shift is visualized (data not shown). In contrast, when gentamicin is added after the renaturation step has proceeded to 75°C, a mobility shift is not produced. These results are consistent with the notion that gentamicin may recognize and interact with an RNA structure formed early in the renaturation process.

8. EXAMPLE: IDENTIFICATION OF A DYE-LABELED TARGET RNA BOUND TO SMALL MOLECULAR WEIGHT COMPOUNDS BY CAPILLARY ELECTROPHORESIS

The results presented in this Example indicate that interactions between a peptide and its target RNA, such as the Tat peptide and TAR RNA, can be monitored by gel retardation assays in an automated capillary electrophoresis system.

8.1. Materials and Methods

8.1.1. Buffers

Tris-potassium chloride (TK) buffer is composed of 50 mM Tris-HCl pH 7.4, 20mM KCl, 0.1%Triton X-100, and 0.5mM $MgCl_2$. Tris-borate-EDTA (TBE) buffer is composed of 45 mM Tris-borate pH 8.0, and 1 mM EDTA. Tris-Potassium chloride-magnesium (TKM) buffer is composed of 50 mM Tris-HCl pH 7.4, 20mM KCl, 0.1%Triton X-100 and 5mM $MgCl_2$.

8.1.1. Gel Retardation Analysis Using Capillary Electrophoresis

RNA oligonucleotides were purchased from Dharmacon, Inc, Lafayette, CO). 500 pmole of a 5' fluorescein labeled oligonucleotide corresponding to the HIV-1 TAR element TAR RNA (5'-GGCGUCACACCUUCGGGUGAAGUCGCC-3' (SEQ ID NO: 30); Huq *et al.*, 1999, Nucleic Acids Research. 27:1084-1093; Hwang *et al.*, 1999, Proc. Natl. Acad. Sci. USA 96:12997-13002) was used. For Tat-TAR gel retardation reactions the method of Huq *et al.* (Nucleic Acids Research, 1999, 27:1084-1093) was

utilized with TK buffer containing 0.5mM MgCl₂ and a 12-mer Tat peptide (YGRKKRRQRRRP (SEQ ID NO: 31); single letter amino acid code). In 20 µl reaction volumes 50 pmoles of labeled oligonucleotide and the short Tat peptide (Tat₄₇₋₅₈) in TK or TKM buffer were heated at 90°C for 2 minutes and allow to cool to room temperature (approximately 24°C) over 2 hours. The reactions were loaded onto a SCE9610 automated capillary electrophoresis apparatus (SpectruMedix; State College, Pennsylvania).

8.2. Results

As presented in the previous Example in Section 7, interactions between a peptide and RNA can be monitored by gel retardation assays. It was hypothesized that interactions between a peptide and RNA could be monitored by gel retardation assays by an automated capillary electrophoresis system. To test this hypothesis, a gel retardation assay by an automated capillary electrophoresis system was performed using the Tat₄₇₋₅₈ peptide and the TAR RNA oligonucleotide. As shown in Figure 5 using the capillary electrophoresis system, in the presence of the Tat peptide, a clear shift is visible upon the addition of increasing concentrations of Tat peptide. In the reaction that lacks peptide, only a peak corresponding to the free RNA is observed. These observations confirm previous reports made using other Tat peptides (Hamy *et al.*, 1997, Proc. Natl. Acad. Sci. USA 94:3548-3553; Huq *et al.*, 1999, Nucleic Acids Res. 27: 1084-1093).

The present invention is not to be limited in scope by the specific embodiments described herein. Indeed, various modifications of the invention in addition to those described will become apparent to those skilled in the art from the foregoing description and accompanying figures. Such modifications are intended to fall within the scope of the appended claims.

Various publications are cited herein, the disclosures of which are incorporated by reference in their entireties.

The invention can be illustrated by the following embodiments enumerated in the numbered paragraphs that follow:

- 5 1. A method for identifying a test compound that binds to a target RNA molecule, comprising the steps of (a) contacting a detectably labeled target RNA molecule with a library of test compounds under conditions that permit direct binding of the labeled target RNA to a member of the library of test compounds so that a detectably labeled target RNA:test compound complex is formed; (b) separating the detectably labeled target
10 RNA:test compound complex formed in step(a) from uncomplexed target RNA molecules and test compounds; and (c) determining a structure of the test compound bound to the RNA in the RNA:test compound complex.
- 15 2. The method of paragraph 1 in which the target RNA molecule contains an HIV TAR element, internal ribosome entry site, "slippery site", instability element, or adenylate uridylate-rich element.
- 20 3. The method of paragraph 1 in which the RNA molecule is an element derived from the mRNA for tumor necrosis factor alpha ("TNF- α "), granulocyte-macrophage colony stimulating factor ("GM-CSF"), interleukin 2 ("IL-2"), interleukin 6 ("IL-6"), vascular endothelial growth factor ("VEGF"), human immunodeficiency virus I ("HIV-1"), hepatitis C virus ("HCV" - genotypes 1a & 1b), ribonuclease P RNA ("RNaseP"), X-linked inhibitor of apoptosis protein ("XIAP"), or survivin.
- 25 4. The method of paragraph 1 in which the detectably labeled RNA is labeled with a fluorescent dye, phosphorescent dye, ultraviolet dye, infrared dye, visible dye, radiolabel, enzyme, spectroscopic colorimetric label, affinity tag, or nanoparticle.
- 30 5. The method of paragraph 1 in which the test compound is selected from a combinatorial library comprising peptoids; random bio-oligomers; diversomers such as hydantoins, benzodiazepines and dipeptides; vinylogous polypeptides; nonpeptidal peptidomimetics; oligocarbamates; peptidyl phosphonates; peptide nucleic acid libraries; antibody libraries; carbohydrate libraries; and small organic molecule libraries, including but not limited to, libraries of benzodiazepines, isoprenoids, thiazolidinones,
35 metathiazanones, pyrrolidines, morpholino compounds, or diazepindiones.

6. The method of paragraph 1 in which screening a library of test compounds comprises contacting the test compound with the target nucleic acid in the presence of an aqueous solution, the aqueous solution comprising a buffer and a combination of salts, preferably approximating or mimicking physiologic conditions.
7. The method of paragraph 6 in which the aqueous solution optionally further comprises non-specific nucleic acids comprising DNA, yeast tRNA, salmon sperm DNA, homoribopolymers, and nonspecific RNAs.
8. The method of paragraph 6 in which the aqueous solution further comprises a buffer, a combination of salts, and optionally, a detergent or a surfactant. In another embodiment, the aqueous solution further comprises a combination of salts, from about 0 mM to about 100 mM KCl, from about 0 mM to about 1 M NaCl, and from about 0 mM to about 200 mM MgCl₂. In a preferred embodiment, the combination of salts is about 100 mM KCl, 500 mM NaCl, and 10 mM MgCl₂. In another embodiment, the solution optionally comprises from about 0.01% to about 0.5% (w/v) of a detergent or a surfactant.
9. Any method that detects an altered physical property of a target nucleic acid complexed to a test compound from the unbound target nucleic acid may be used for separation of the complexed and non-complexed target nucleic acids in the method of paragraph 1. In a preferred embodiment, electrophoresis is used for separation of the complexed and non-complexed target nucleic acids. In a preferred embodiment, the electrophoresis is capillary electrophoresis. In other embodiments, fluorescence spectroscopy, surface plasmon resonance, mass spectrometry, scintillation, proximity assay, structure-activity relationships ("SAR") by NMR spectroscopy, size exclusion chromatography, affinity chromatography, and nanoparticle aggregation are used for the separation of the complexed and non-complexed target nucleic acids.
10. The structure of the test compound of the RNA: test compound complex of paragraph 1 is determined, in part, by the type of library of test compounds. In a preferred embodiment wherein the combinatorial libraries are small organic molecule libraries, mass spectroscopy, NMR, or vibration spectroscopy are used to determine the structure of the test compounds.

WHAT IS CLAIMED IS:

1. A method for identifying a test compound that binds to a target RNA
5 molecule, comprising the steps of:
- (a) contacting a detectably labeled target RNA molecule with a
library of test compounds under conditions that permit direct
binding of the labeled target RNA to a member of the library
of test compounds so that a detectably labeled target
10 RNA:test compound complex is formed;
 - (b) separating the detectably labeled target RNA:test compound
complex formed in step(a) from uncomplexed target RNA
molecules and test compounds by capillary gel
electrophoresis; and
 - (c) 15 determining a structure of the test compound bound to the
RNA in the RNA:test compound complex by mass
spectroscopy.

20

25

30

35

AMENDED CLAIMS

[received by the International Bureau on 17 September 2002 (17.09.02);
Claims 1 to 10 replaced by new claims 1 to 19. (3 sheets)]

- 5 1. A method for identifying a test compound that binds to a target RNA molecule, comprising the steps of:
- (a) contacting a detectably labeled target RNA molecule with a library of test compounds under conditions that permit direct binding of the labeled target RNA to a member of the library of test compounds so that a detectably labeled target RNA:test compound complex is formed;
- 10 (b) separating the detectably labeled target RNA:test compound complex formed in step (a) from uncomplexed target RNA molecules and test compounds; and
- 15 (c) determining a structure of the test compound bound to the RNA in the RNA:test compound complex.
2. The method of claim 1 in which the target RNA molecule contains an HIV TAR element, internal ribosome entry site, "slippery site", instability element, or
- 20 adenylate uridylate-rich element.
3. The method of claim 1 in which the RNA molecule is an element derived from the mRNA for tumor necrosis factor alpha ("TNF- α "), granulocyte-macrophage colony stimulating factor ("GM-CSF"), interleukin 2 ("IL-2"), interleukin 6
- 25 ("IL-6"), vascular endothelial growth factor ("VEGF"), human immunodeficiency virus I ("HIV-1"), hepatitis C virus ("HCV" - genotypes 1a & 1b), ribonuclease P RNA ("RNaseP"), X-linked inhibitor of apoptosis protein ("XIAP"), or survivin.
4. The method of claim 1 in which the detectably labeled RNA is
- 30 labeled with a fluorescent dye, phosphorescent dye, ultraviolet dye, infrared dye, visible dye, radiolabel, enzyme, spectroscopic colorimetric label, affinity tag, or nanoparticle.
5. The method of claim 1 in which the test compound is selected from a combinatorial library comprising peptoids; random bio-oligomers; diversomers such as
- 35 hydantoin, benzodiazepines and dipeptides; vinyllogous polypeptides; nonpeptidal

peptidomimetics; oligocarbamates; peptidyl phosphonates; peptide nucleic acid libraries; antibody libraries; carbohydrate libraries; or small organic molecule libraries.

5

6. The method of claim 5 in which the small organic molecule libraries are libraries of benzodiazepines, isoprenoids, thiazolidinones, metathiazanones, pyrrolidines, morpholino compounds, or diazepindiones.

10

7. The method of claim 1 in which screening a library of test compounds comprises contacting the test compound with the target nucleic acid in the presence of an aqueous solution wherein the aqueous solution comprises a buffer and a combination of salts.

15

8. The method of claim 7 wherein the aqueous solution approximates or mimics physiologic conditions.

9. The method of claim 7 in which the aqueous solution optionally further comprises non-specific nucleic acids comprising DNA, yeast tRNA, salmon sperm DNA, homoribopolymers, and nonspecific RNAs.

20

10. The method of claim 7 in which the aqueous solution further comprises a buffer, a combination of salts, and optionally, a detergent or a surfactant.

25

11. The method of claim 10 in which the aqueous solution further comprises a combination of salts, from about 0 mM to about 100 mM KCl, from about 0 mM to about 1 M NaCl, and from about 0 mM to about 200 mM MgCl₂.

12. The method of claim 11 wherein the combination of salts is about 100 mM KCl, 500 mM NaCl, and 10 mM MgCl₂.

30

13. The method of claim 10 wherein the solution optionally comprises from about 0.01% to about 0.5% (w/v) of a detergent or a surfactant.

35

15. The method of claim 1 in which separating the detectably labeled target RNA:test compound complex formed in step (a) from uncomplexed target RNA and
5 test compounds is by electrophoresis.
16. The method of claim 15 in which the electrophoresis is capillary electrophoresis.
- 10 17. The method of claim 1 in which separating the detectably labeled target RNA:test compound complex formed in step (a) from uncomplexed target RNA and test compounds is by fluorescence spectroscopy, surface plasmon resonance, mass spectrometry, scintillation, proximity assay, structure-activity relationships ("SAR") by NMR spectroscopy, size exclusion chromatography, affinity chromatography, or
15 nanoparticle aggregation.
18. The method of claim 1 in which the library of test compounds are small organic molecule libraries.
- 20 19. The method of claim 18 in which the structure of the test compound is determined by mass spectroscopy, NMR, or vibration spectroscopy.

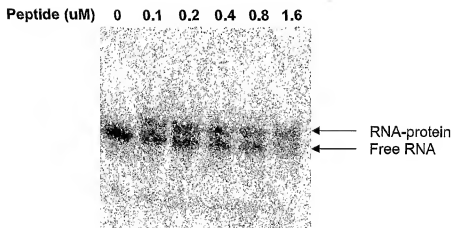
Figure 1**Sheet 1/5****Attorney Docket No. 10589-007**

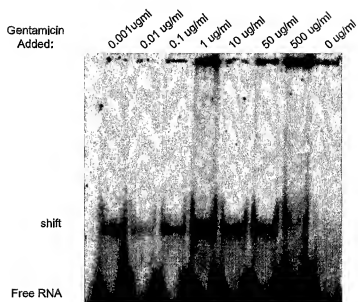
Figure 2**Sheet 2/5****Attorney Docket No. 10589-007**

Figure 3

Sheet 3/5

Attorney Docket No. 10589-007

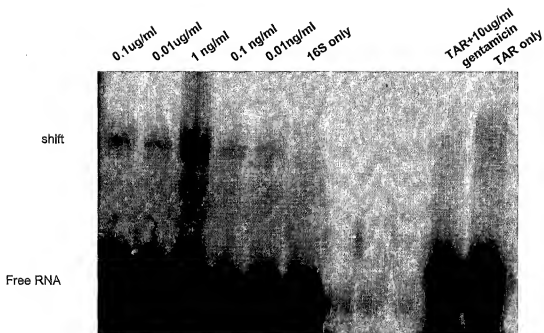


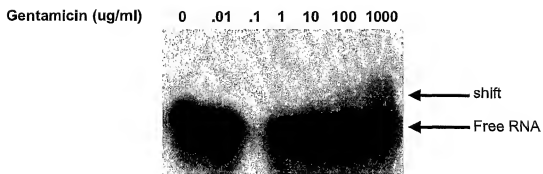
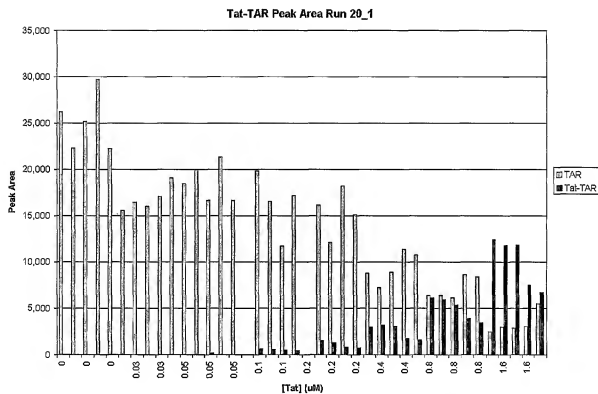
Figure 4**Sheet 4/5****Attorney Docket No. 10589-007**

Figure 5
Sheet 5/5
Attorney Docket No. 10589-007



SEQUENCE LISTING

<110> PTC Therapeutics, Inc.

<120> METHODS FOR IDENTIFYING SMALL MOLECULES THAT BIND SPECIFIC RNA
STRUCTURAL MOTIFS

<130> 10589-007-228

<140> To be assigned

<141> 2002-04-11

<150> 60/282,965

<151> 2001-04-11

<160> 31

<170> PatentIn version 3.0

 $\langle 210 \rangle$ 1

<211> 21

<212> RNA

<213> Homo sapiens

```
<400> 1
uuuuuuuuuu uuuuuuuuuu a
```

<210> 2

<211> 17

<212> RNA

<213> Homo sapiens

```
<400> 2  
auuuuuuuuu uuauuuu
```

<210> 3
 <211> 15
 <212> RNA
 <213> Homo sapiens

<400> 3
 wauuuuuuuu uuuw 15

<210> 4
 <211> 13
 <212> RNA
 <213> Homo sapiens

<400> 4
 wwauuuuuuu aww 13

<210> 5
 <211> 13
 <212> RNA
 <213> Homo sapiens

<400> 5
 wwwauuuuuu www 13

<210> 6
 <211> 1643
 <212> DNA
 <213> Homo sapiens

<400> 6
 gcagaggacc agctaagagg gagagaagca actacagacc ccccttgaaa acaaccctca 60
 gacgccacat cccctgacaa gctgccaggc aggttctctt cctctcacat actgaccac 120
 ggotccaccc tctctccct ggaaaggaca ccatgagcac tgaagcatg atccgggacg 180
 tggagctggc cgaggaggcg ctccccaaga agacaggggg gccccagggc tccaggcggt 240
 gcttgttctc cagcctcttc tcttctctga tcgtggcagg cgccaccacg ctcttctgcc 300

tgctgcactt tggagtgtc ggcgccaga ggaagagtt cccagggac ctctctctaa	360
tcagccctct ggcocaggca gtcagatcat ctctcgaac cccgagtgc aagcctgtag	420
cccatgttgt agcaaacct caagctgagg ggcagctcca gtggctgaac cgcggggcca	480
atgcctctct ggcacatggc gtggagctga gagataacca gctgtgtgtg ccatcagagg	540
gcctgtacct catctactcc caggtctctc tcaagggcca aggtctcccc tccacccatg	600
tgctctctac ccacaccatc agcgcgatcg ccgtctccta ccagaccaag gtcaacctcc	660
ctctctccat caagagcccc tgccagaggg agaccccaga gggggctgag gccaaagcct	720
ggtatgagcc catctatctg ggaggggtct tccagctgga gaagggtgac cgaactcagc	780
ctgagatcaa tcggcccgac tatctcgact ttgcagagtc tgggcaggtc tactttggga	840
tcattgccct gtgaggagga cgaacatcca acctcccaa acgcctcccc tgccccaatc	900
ccctttattac ccctccttc agacaccctc aacctcttct ggctcaaaaa gagaattggg	960
ggcttagggc cgaacccaa gcttagaact ttaagcaaca agaccaccac ttogaacact	1020
gggattcagg aatgtgtggc ctgcacagtg aattgctggc aaccactaag aattcaaat	1080
ggggcctcca gaactcactg gggcctacag ctttgatccc tgacatctg aatctggaga	1140
ccaggagacc tttgtttctg gccagaatgc tgcaggactt gagaagacct cacctagaaa	1200
ttgacacaag tggaccttag gccttctctc ctccagatgt ttccagactt ccttgagaca	1260
cgagagccag ccctcccat ggagccagct cctctatatt atgtttgcac ttgtgattat	1320
ttattattta ttattatttt atttatttac agatgaatgt atttatttgg gagaccgggg	1380
tatcctgggg gaccacatgt aggagctgcc ttggctcaga catgttttcc gtgaaaacgg	1440
agctgaacaa taggctgttc coactgtagc ccctggcctc tgtgccttct ttgattatg	1500
ttttttaaaa tatttatctg attaagttgt ctaaacaatg ctgatttggt gaccaactgt	1560
cactcattgc tgagcctctg ctcccaggg gaggttgtgc tgtaaatgcc ctactattca	1620
gtggcgagaa ataaagtttg ctt	1643

<210> 7

<211> 756

<212> DNA

<213> Homo sapiens

<400> 7

gctggaggat gtggctgcag agcctgctgc tcttgggcac tgtggcctgc agcatctctg	60
caccgcccg ctgcgccagc cccagcacgc agccctggga gcactgtgaat gccatccagg	120
agggccggcg tctcctgaac ctgagtagag aactgctgc tgagatgaat gaaacagtag	180

```

aagtcatttc agaaatgttt gacctccagg agccgacctg cctacagacc cgccctggagc 240
tgtacaagca gggcctgcgg ggcagcctca ccaagctcaa gggcccccctg accatgatgg 300
ccagccacta caagcagcac tgcctccaa ccccggaac ttccgtgtga acccagacta 360
tcacctttga aagtttcaa gagaacctga aggaatttct gcttgtcatc ccctttgact 420
gctggggacc agtcaccagg tgagaccggc cagatgaggc tggccaagcc ggggagctgc 480
tctctcatga aacaagagct agaaactcag gatggtcatc ttggagggac caaggggtgg 540
gccacagcca tgggtggagt ggccctggacc tgcctgggc cactactgacc ctgatacagg 600
catggcgcaa gaatgggaat attttatact gacagaaatc agtaatatatt atatatattat 660
atttttaaaa tattttatta tttatttatt taagttcata ttccatatatt attcaagatg 720
ttttaccgta ataattatta ttaaaaatat gcttct 756

```

<210> 8

<211> 756

<212> DNA

<213> Homo sapiens

```

<400> 8
tctgggggat gtggctgcag agcctgctgc tcttgggcac tgtggcctgc agcatctctg 60
caccgcccg ctcgccagc cccagcacgc agccctggga gcattgtgaat gccatccagg 120
aggcccgccg tctctgaac ctgagtagag aactgtctgc tgagatgaat gaaacagtag 180
aagtcatttc agaaatgttt gacctccagg agccgacctg cctacagacc cgccctggagc 240
tgtacaagca gggcctgcgg ggcagcctca ccaagctcaa gggcccccctg accatgatgg 300
ccagccacta caagcagcac tgcctccaa ccccggaac ttccgtgtga acccagacta 360
tcacctttga aagtttcaa gagaacctga aggaatttct gcttgtcatc ccctttgact 420
gctggggacc agtcaccagg tgagaccggc cagatgaggc tggccaagcc ggggagctgc 480
tctctcatga aacaagagct agaaactcag gatggtcatc ttggagggac caaggggtgg 540
gccacagcca tgggtggagt ggccctggacc tgcctgggc cactactgacc ctgatacagg 600
catggcgcaa gaatgggaat attttatact gacagaaatc agtaatatatt atatatattat 660
atttttaaaa tattttatta tttatttatt taagttcata ttccatatatt attcaagatg 720
ttttaccgta ataattatta ttaaaaatat gcttct 756

```

<210> 9

<211> 825

<212> DNA

<213> Homo sapiens

```

<400>  9
atcactctct ttaatcacta ctcacattaa cctcaactcc tgccacaatg tacaggatgc      60
aactcctgtc ttgcattgca ctaattcttg cacttgtcac aaacagtgca cctacttcaa    120
gttcgacaaa gaaaacaaag aaaacacagc tacaactgga gcatttactg ctggatttac    180
agatgatttt gaatggaatt aataattaca agaatcccaa actcaccagg atgtcacat      240
ttaagtttta catgcccaag aaggccacag aactgaaaca gcttcagtgt ctagaagaag    300
aactcaaac cctggaggaa gtgctgaatt tagctcaaag caaaaacttt cacttaagac    360
ccagggaact aatcagcaat atcaacgtaa tagttctgga actaaagga tctgaacaa      420
cattcatgtg tgaatatgca gatgagacag caaccattgt agaatttctg aacagatgga    480
ttacottttg tcaaaagcat atctcaacac taacttgata attaatgtct tccacttaa      540
aacatatcag gccttctatt tatttattta aatattttaa ttttatattt attgttgaat    600
gtatggttgc tacottattg aactattatt cttaacttta aaactataaa tatggatctt    660
ttatgattct ttttgtaagc cctaggggct ctaaaatggt ttacottatt tatocccaaa      720
atatttatta ttatgttgaa tgttaaatat agtatctatg tagattgggt agtaaaacta      780
ttaataaat ttgataataa taaaaaaaaa aacaaaaaaa aaaaa                                825

```

<210> 10

<211> 15

<212> RNA

<213> Homo sapiens

<220>

<221> misc_feature

<222> (1)..(1)

<223> N = A, U, G, OR C

<220>

<221> misc_feature

<222> (15)..(15)

<223> N = A, U, G, OR C

<400> 10
 nauuuuuuuu uuuuu 15

<210> 11

<211> 1125

<212> DNA

<213> Homo sapiens

<400> 11
 ttctgcccctc gagcccaaccg ggaacgaaag agaagctcta tctgcctcc aggagcccag 60
 ctatgaactc cttctccaca agcgcccttcg gtccagttgc cttctccctg gggctgctcc 120
 tgggtgttgcc tgetgccttc cctgccccag taaccccagg agaagattcc aaagatgtag 180
 ccgccccaca cagacagcca ctcacctctt cagaacgaat tgacaaacaa attcgtgtaca 240
 tctctcagcg catctcagcc ctgagaaagg agacatgtaa caagagtaac atgtgtgaaa 300
 gcagcaaaaga ggcactggca gaaaacaacc tgaaccttcc aaagatggct gaaaaagatg 360
 gatgcttcca atctggattc aatgaggaga cttgcctggt gaaaatcatc actggtcttt 420
 tggagtttga ggtataccta gagtacctcc agaacagatt tgagagtagt gaggaacaag 480
 ccagagctgt gcagatgagt acaaaagtcc tgatccagtt cctgcagaaa aaggcaaaaga 540
 atctagatgc aataaccacc cctgacccaa ccacaaatgc cagcctgctg acgaagctgc 600
 aggcacagaa ccagtggtg caggacatga caactcatct cattctgcgc agctttaagg 660
 agttcttcca gtccagcctg agggctcttc ggcaaatgta gcatgggac ctcagattgt 720
 tgttgttaat gggcattcct tcttctggtc agaaacctgt ccaactggga cagaacttat 780
 gttgttctct atggagaact aaaagtatga gcgttaggac actattttaa ttatttttaa 840
 tttattaata tttaaatatg tgaagctgag ttaattttat taagtcatat ttatattttt 900
 aagaagtacc acttgaacaa ttttatgtat tagttttgaa ataataatgg aaagtggcta 960
 tgcagtttga atatcctttg ttccagagcc agatcatttc ttggaaagtg taggcttacc 1020
 tcaataaatg ggctaactta tacatatatt taaagaaata ttatatattg atttatataa 1080
 tgtataaatg gttttttata caataaatgg cattttaaaa aattc 1125

<210> 12

<211> 3166

<212> DNA

<213> Homo sapiens

<400> 12
 aagagctcca gagagaagtc gaggaagaga gagacggggt cagagagagc gcgcgggcgt 60
 gcgagcagcg aaagcgacag gggcaagtg agtgacctgc ttttgggggt gaccgcggga 120
 gcgcggcggtg agccctcccc cttgggatcc cgcagctgac cagtgcgcgt gacggacaga 180
 cagacagaca ccgccccag cccagttac caccctctcc ccggccggcg gcggacagtg 240
 gacgcggcgcg cgagccgcgg gcaggggcgg gagcccgccc ccggaggcgg ggtggagggg 300
 gtccggagctc gcggcgctgc actgaaaactt ttctgtccaac ttctgggctg ttctcgcttc 360
 ggaggagcgg tgggtccgcg gggggaagcc gagccgagcg gagccgcgag aagtgtctagc 420
 tcgggcccgg aggagccgca gccggaggag ggggaggagg aagaagagaa ggaaggagg 480
 agggggccgc agtgcgact cggcgctcgg aagccgggct catggaaggg tgaggcggcg 540
 gtgtgcgcag acagtgtccc agcgcgcgcg ctcccagcc ctggcccgcg ctccggccgg 600
 gaggaagagt agctcgccga ggcgcgcgag agagcgggcc gccccacagc ccgagccgga 660
 gagggacgcg agccgcgcgc cccggtcggg cctccgaaac catgaacttt ctgctgtctt 720
 ggggtgcatlg gagccttgcc ttgctgtctt acctccacca tgccaagtgg tcccaggctg 780
 ccccattggc agaaggagga gggcagaatc atcacgaagt ggtgaagtgc atggatgtct 840
 atcagcgcag ctactgcat ccaatogaga cctggtgga catottccag gattacctg 900
 atgagatcga gtacatcttc aagccatcct gtgtgccctt gatgcgatgc gggggctgct 960
 ccaatgacga gggcctggag tgtgtgccca ctgaggagtc caacatcacc atgcagatta 1020
 tgcggatcaa acctcaccaa ggccagcaca taggagagat gagcttccca cagcacaaca 1080
 aatgtgaatg cagaccaag aaagatagag caagacaaga aaatccctgt gggccttgct 1140
 cagagcggag aaagcatttg tttgtacaag atccgcagac gtgtaaatgt tcttgcaaaa 1200
 acacacactc gcgttgcaag gcgaggcagc ttgagttaaa cgaacgtact tgcagatgtg 1260
 acaagccgag gcggtgagcc gggcaggagg aaggagcctc cctcagggtt tcgggaacca 1320
 gatctctctc caggaaagac tgatacagaa cgatcgatac agaaaccacg ctgcgcgcac 1380
 cacaccatca ccatcgacag aacagtcctt aatccagaaa cctgaaatga aggaagagga 1440
 gactctgcgc agagcacttt gggtcgggag ggcgagactc cggcgggaagc attccggggc 1500
 ggggtgaccca gcacggtccc tcttggaatt ggattogcca ttttattttt cttgotgcta 1560
 aatcacccag ccgggaagat tagagagttt tttttctggg attcctgtag acacaccac 1620
 ccacatacat acatttatat atatatatata tatatatata taaaaataaa tatctctatt 1680
 ttatatatat aaaatatata tattcttttt ttaaattaac agtgctaagt ttattggtgt 1740
 ctccactgga tgtatttgac tgctgtggac ttgagttggg aggggaatgt tccactcag 1800

atcctgacag	ggaagaggag	gagatgagag	actctggcat	gatctttttt	ttgtcccaact	1860
tggtggggcc	agggtcctct	cccctgccca	agaatgtgca	aggccagggc	atgggggcaa	1920
atatgaccca	gttttgggaa	cacgcacaaa	cccagccctg	gcgctgagcc	tctctacccc	1980
aggtoacagc	gacagaaaga	caaatacacg	gttcggggat	gaggacacgc	gctctgacca	2040
ggagtittgg	gagcttcagg	acattgctgt	gctttgggga	ttccctccac	atgtctgacg	2100
cgcatctcgc	ccccaggggc	actgcctgga	agattcagga	gcctggggcg	ccttgcgtta	2160
ctctcacctg	cttctgagtt	gcccaggagg	ccactggcag	atgtcccgcc	gaagagaaga	2220
gacacattgt	tggaagaagc	agcccatgac	agcgccctt	cctgggactc	gcctccatcc	2280
tcttctctgt	ccccttccctg	gggtgcagcc	taaaaggacc	tatgtctcta	caccattgaa	2340
accactagtt	ctgtccccc	aggaaacctg	gttgtgtgtg	tgtgagtgtg	tgaccttct	2400
ccatccctg	gtccttccct	tccctcccg	aggcacagag	agacagggca	ggatccacgt	2460
gcccattgtg	gaggcagaga	aaagagaaag	tgttttatat	acggctacta	tttaatatcc	2520
ctttttaatt	agaaataga	acagttaatt	taattaaaga	gtagggtttt	tttctagtat	2580
tcttggttaa	tatttaattt	caactattta	tgagatgtat	cttttgcctc	ctcttgcctc	2640
cttatttgta	ccggtttttg	tataataaat	tcattgtttc	aatctctctc	tccttgatcg	2700
gtgacagtca	ctagcttctc	ttgaacagat	atttaatttt	gctaaccactc	agctctgccc	2760
tcccgcgtcc	cctggctccc	cagcacacat	tcctttgaaa	gagggtttca	atatacatct	2820
acatactata	tatatatttg	gcaacttgta	tttgtgtgta	tatatatata	tatatgttta	2880
tgtatatatg	tgtacctgaa	aaaataaaca	tcgctattct	gttttttata	tgttcaaacc	2940
aaacaagaaa	aaatagagaa	ttctacatac	taaatctctc	tcctttttta	attttaatat	3000
ttgttatcat	ttattttattg	gtgctactgt	ttatccgtaa	taattglggg	gaaaagatat	3060
taacatcacg	tctttgtctc	tagtgcagtt	tttcgagata	ttccgtagta	catattttatt	3120
tttaacaac	gacaagaaa	tacagatata	tcctaaaaaa	aaaaaa		3166

<210> 13

<211> 249

<212> RNA

<213> Homo sapiens

<400> 13

ccgggcucac	ggagggguga	ggcggcggu	ugcgagaca	gugcuccagc	gcgcgcgcuc	60
cccagccucg	gcccggccuc	ggcggcggu	gaagaguagc	ucgcccaggc	gccgaggaga	120
gcggggcgcc	ccacagcccg	agccggagag	ggacgcgagc	cgcgcgcccc	ggucggggccu	180

ccgaaaccuu gaacuuuucug cugucuuuggg ugcauuggag ccuugccuug cugcucuacc 240
uccaccaug 249

<210> 14

<211> 9181

<212> DNA

<213> Homo sapiens

<400> 14
ggtctctctg gttagaccag atctgagcct gggagctctc tggctaacta gggaaccac 60
tgcttaagcc tcaataaagc ttgccttgag tgcctcaagt agtgtgtgcc cgtctgttgt 120
gtgactctgg taactagaga tccctcagac ccttttagtc agtgtggaaa atctctagca 180
gtggcgcccg aacagggacc tgaagcgaa agggaaacca gaggagctct ctgcagcgag 240
gactcggett gctgaagcgc gcacggcaag agggaggggg cggcgactgg tgaatcgcc 300
aaaaattttg actagcggag gctagaagga gagagatggg tgcgagagcg tcagtattaa 360
cgcggggaga attagatcga tgggaaaaaa ttcggttaag gccaggggga aagaaaaaat 420
ataaattaaa acatatagta tgggcaagca gggagctaga acgattcgca gttaatcctg 480
gcctgttaga aacatcagaa ggctgtagac aaatactggg acagctacaa ccatcccttc 540
agacaggatc agaagaactt agatcattat ataatacagt agcaaccctc tatttgtgtc 600
atcaaaggat agagataaaa gacaccaagg aagctttaga caagatagag gaagagcaaa 660
acaaaagtaa gaaaaagca cagcaagcag cagctgacac aggacacagc aatcaggta 720
gccaaaatta ccctatagtg cagaacatcc aggggcaaat ggtacatcag gccatatcac 780
ctagaacttt aaatgcattg gtaaaagtag tagaagagaa ggctttcagc ccagaagtga 840
taccatggtt ttacgacatta tcagaaggag ccacccaca agatttaaac accatgctaa 900
acacagtggt gggacatcaa gcagccatgc aaatgttaaa agagaccatc aatgaggaa 960
ctgcagaatg gtagatagtg catccatgct atgcagggcc tattgcacca gccagatga 1020
gagaaccaag gggaagtgc atagcaggaa ctactagtac ccttcaggaa caaataggat 1080
ggatgacaaa taatccact atccagtag gagaaattta taaaagtgg ataactcctg 1140
gattaaataa aatagtaaga atgtatagcc ctaccagcat tctggacata agacaaggac 1200
caaaggaacc ctttagagac tatgtagacc ggttctataa aactctaaga gcgagcaag 1260
cttcacagga ggtaaaaaat tggatgacag aaacottgtt ggtccaaaat gcgaaccag 1320
attgtaagac tatttttaaa gcattgggac cagoggctac actagaagaa atgatgacag 1380
catgtcaggg agtaggagga ccgggcata aggaagagt ttggctgaa gcaatgagcc 1440

aagtaacaaa ttcagctacc ataatgatgc agagaggcaa ttttaggaac caaagaaaga	1500
ttgttaagtg tttoaatgtg ggcaaaagaag ggcacacagc cagaaattgc agggccoccta	1560
ggaaaaaggg ctgttggaag tgtggaaagg aaggacacca aatgaaagat tgtaactgaga	1620
gacaggctaa ttttttaggg aagatctggc cttcctacaa gggaaggcca gggaatttct	1680
ttcagagcag accagagcca acagcccccac cagaagagag cttcaggtct ggggtgaga	1740
caacaactcc ccttcagaag caggagccga tagacaagga actgtatcct ttaacttccc	1800
tcaggtoact ctttgccaac gacccctcgt cacaataaag ataggggggc aactaaagga	1860
agctctatta gatacaggag cagatgatac agtattagaa gaaatgagtt tgccagggaag	1920
atggaacca aaaaatgatag ggggaattgg aggttttata aaagtaagac agtatgatca	1980
gataactcata gaaatctgtg gacataaagc tataggtaac gtattagtag gacctacacc	2040
tgtaacata attggaagaa atctgttgac tcagattggg tgcaacttaa attttcccat	2100
tagccctatt gagactgtac cagtaaaatt aaagccagga atgtagggcc caaaagttaa	2160
acaatggcca ttgacagaag aaaaaataaa agcattagta gaaatttgta cagagatgga	2220
aaaggaaggg aaaatttcaa aaattgggcc tgaaaatcca tacaatactc cagtatttgc	2280
cataaagaaa aaagacagta ctaaatggag aaaattagta gatttcagag aacttaataa	2340
gagaactcaa gacttctggg aagtccaatt aggaatacca catcccgagc gggtaaaaaa	2400
gaaaaatca gtaacagtac tggatgtggg tgatgcata ttttcagttc ccttagatga	2460
agacttcagg aagtatactg catttaccat acctagtata acaatgaga caaccgggat	2520
tagatatcag tacaatgtgc ttccacaggg atggaagga tcaccagcaa tattccaaag	2580
tagcatgaca aaaaacttag agccttttag aaaacaaaat ccagacatag ttatctatca	2640
atacatggat gatttgtatg taggatctga cttagaaata gggcagcata gaacaaaaat	2700
agaggagctg agacaacatc tgttgagggt gggacttacc acaccagaca aaaaacatca	2760
gaaagaacct ccattccttt ggatgggcta tgaactccat cctgataaat ggacagtaca	2820
gcctatagtg ctgccagaaa aagacagctg gactgtcaat gacatacaga agttagtggg	2880
gaaattgaat tgggcaagtc agatttacc agggattaaa gtaaggcaat tatgtaaact	2940
ccttagagga accaaagcac taacagaagt aataccacta acagaagaag cagagctaga	3000
actggcgaaa aacagagaga ttctaaaaga accagtacat ggagtgtatt atgaccatc	3060
aaaagactta atagcagaaa tacagaagca ggggcaaggc caatggacat atcaaaattta	3120
tcaagagcca tttaaaaatc tgaaaacagg aaaatatgca agaattgagg gtgccccac	3180
taatgatgta aaacaattaa cagaggcagt gcaaaaaata accacagaaa gcatagtaat	3240
atggggaag actcctaact ttaactgcc catacaaaag gaaacatggg aaacatggtg	3300
gcagagtat tggcaagcca cctggattcc tgagtgggag tttgttaata cccctccctt	3360

agtgaatta tggtagcag	tagagaaaga	acccatagta	ggagcagaaa	ccttctatgt	3420
agatggggca gctaacagg	agactaaatt	aggaaaagca	ggatatgtta	ctaatagagg	3480
aagacaaaaa gttgtcacc	taactgacac	aacaaatcag	aagactgagt	tacaagcaat	3540
ttatctagct ttgcaggatt	cggtattaga	agtaaacata	gtaacagact	cacaatatgc	3600
attaggaatc attcaagcac	aaccagatca	aagtgaatca	gagttagtca	atcaaaatat	3660
agagcagtta ataaaaaagg	aaaagggtcta	tctggcatgg	gtaccagcac	acaaagggaat	3720
tggaggaaat gaacaagtag	ataaattagt	cagtgcctga	atcaggaaag	tactattttt	3780
agatggaata gataaggccc	aagatgaaca	tgagaaatat	cacagtaatt	ggagagcaat	3840
ggctagtgat tttaacctgc	cacctgtagt	agcaaaagaa	atagttagcca	gctgtgataa	3900
atgtcagcta aaaggagaag	ccatgcatgg	acaagtagac	tgtagtccag	gaatatggca	3960
actagattgt acacatttag	aaggaaaagt	tatcctggta	gcagtctcat	tagccagtgg	4020
atatatagaa gcagaagtta	ttccagcaga	aacagggcag	gaaacagcat	attttctttt	4080
aaaattagca ggaagatggc	cagtaaaaac	aatacatact	gacaatggca	gcaatttcac	4140
cggtgctacg gttatggccg	cctgttggtg	ggcgggaatc	aagcaggaat	ttggaattcc	4200
ctacaatccc caaagtcaag	gagtagtaga	atctatgaat	aaagaattaa	agaaaattat	4260
aggcagagta agagatcagg	ctgaacatct	taagacagca	gtacaatagg	cagtattcat	4320
ccacaatttt aaaaagaaa	gggggattgg	gggtacagt	gcaggggaaa	gaatagtaga	4380
cataatagca acagacatac	aaactaaaga	attacaaaaa	caaattacaa	aaattcaaaa	4440
ttttcggttt tattacagg	acagcagaaa	tcacttttg	aaaggaccag	caaagctcct	4500
ctggaagggt gaagggcgag	tagtaataca	agataatagt	gacataaaag	tagtgccaag	4560
aagaaaagca aagatcatta	gggattatgg	aaaacagatg	gcagggtgat	attgtgtggc	4620
aagtagacag gatgaggatt	agaacatgga	aaagtttagt	aaaacaccat	atgtatgttt	4680
cagggaaagc taggggatgg	ttttatagac	atcactatga	aagccctcat	ccaagaataa	4740
gttcagaagt acacatccca	ctaggggatg	ctagattggg	aataacaaca	tattggggtc	4800
tgcatacagg agaaagagac	tggcatttgg	gtcagggagt	ctccatagaa	tggaagaaaa	4860
agagatatag cacacaagta	gacctgaac	tagcagacca	actaattcat	ctgtattact	4920
ttgactgttt ttcagactct	gctataagaa	aggccttatt	aggacacata	gttagcccta	4980
gggtgtgaata tcaagcagga	cataacaagg	taggatctct	acaatacttg	gcactagcag	5040
cattaataac accaaaaaag	ataaagccac	ctttgcctag	tgttacgaaa	ctgacagagg	5100
atagatggaa caagcccag	aagaccaagg	gccacagagg	gagccacaca	atgaatggac	5160
actagagcct ttagaggagc	ttaagaatga	agctgttaga	cattttccta	ggatttggct	5220
ccatggccta gggcaacata	totatgaaac	ttatggggat	acttggggag	gagtggaagc	5280

cataataaga attctgcaac aactgctgtt tatccathtt cagaattggg tgtcgacata	5340
gcagaaatagg cgttactcga cagaggagag caagaaatgg agccagtaga tcttagacta	5400
gagccctgga agcatccagg aagtgcagct aaaactgctt gtaccaattg ctattgtaaa	5460
aagtgttgct ttcatgtcca agtttgtttc ataacaaaag ctttaggcat ctccataggc	5520
aggaagaagc ggagacagcg acgaagagct catcagaaca gtcagactca tcaagcttct	5580
ctatcaaagc agtaagtagt acatgtaatg caacctatag caatagtagc aatagtagca	5640
ttagtagtag caataataat agcaatagtt gtgtggtcca tagtaatcat agaataatag	5700
aaaatattaa gacaaagaaa aatagacagg ttaattgata gactaataga aagagcagaa	5760
gacagtggca atgagagtga aggagaaata tcagcacttg tgagatggg ggtggagatg	5820
gggcaccatg ctcttggga tgttgatgat ctgtagtgtc acagaaaaat tgtgggtcac	5880
agtcatttat ggggtacctg tgtggaagga agcaaccacc actctathtt gtgcacaga	5940
tgctaaagca tatgatacag aggtacataa gtgttgggcc acacatgctt gtgtaccac	6000
agaccccaac ccacaagaag tagtattggt aaatgtgaca gaaaatttta acatgtggaa	6060
aaatgacatg gtagaacaga tgcattgagg tataatcagt ttatgggatc aaagcctaaa	6120
gccatgtgta aaattaaccc cactctgtgt tagtttaag tgcactgatt tgaagaatga	6180
tactaatacc aatagtagta ggggagaat gataatggag aaaggagaga taaaaaactg	6240
ctctttcaat atcagcacia gcataagagg taagggtgag aaagaatatg cattttttta	6300
taaacttgat ataataccaa tagataatga tactacoagc tataagttga caagttgtaa	6360
cacctcagtc attacacagg cctgtccaaa ggtatccttt gagccaattc ccatacatta	6420
ttgtgcccg gctggttttg cgaattctaa atgtaataat aagacgttca atggaacagg	6480
accatgtaca aatgtcagca cagtacaatg tacacatgga attaggccag tagtatcaac	6540
tcaactgctg ttaaatggca gtctagcaga agaagaggta gtaattagat ctgtcaatht	6600
cacggacaat gotaaaacca taatagtaca gotgaacaca tctgtagaaa ttaattgtac	6660
aagacccaac aacaatacaa gaaaaagaat ccgtatocag agaggaccag ggagagcatt	6720
tgttacaata ggaaaaatag gaaatatgag acaagcacat tgtaacatta gtatagcaaa	6780
atggaataac actttaaaac agatagctag caaattaaga gaacaatttg gaaataataa	6840
aacaataatc ttttaagcaat cctcaggagg ggaccagaaa attgtaacgc acagttttta	6900
ttgtggaggg gaattttttt actgtaatte aacacaactg ttttaatagta cttggtttta	6960
tagtacttgg agtactgaag ggtcaaataa cactgaagga agtgacacaa tcaacctccc	7020
atgcagaata aaacaaatta taaacatgtg gcagaaagta ggaaagcaa tgtatgccc	7080
tcccatcagt ggacaaatta gatgttcac aaatattaca gggctgctat taacaagaga	7140
tggtggtaat agcaacaatg agtccgagat cttcagacct ggaggaggag atatgaggga	7200

caattggaga agtgaattat ataaatataa agtagtaaaa attgaaccat taggagtagc	7260
acccaccaag gcaaagagaa gagtgggtgca gagagaaaa agagcagtgg gaataggagc	7320
tttgttcctt gggttcttgg gaggcagcagg aagcactatg ggccgagcct caatgacgct	7380
gacggtacag gccagacaat tattgtctgg tatagtgcag cagcagaaca atttgcgtgag	7440
ggctattgag gcgcaacagc atctgttgca actcacagtc tggggcatca agcagctcca	7500
ggcaagaatc ctggctgtgg aaagatacct aaaggatcaa cagctcctgg gggattgggg	7560
ttgctctgga aaactcattt gcaccactgc tgtgccttgg aatgctagtt ggagtaataa	7620
atctctggaa cagatttgga atcacacgac ctggatggag tgggacagag aaattaacaa	7680
ttacacaagc ttaatacact ccttaattga agaatcgcaa aaccagcaag aaaagaatga	7740
acaagaatta ttggaattag ataatgggc aagtttggg aattggttta acataacaaa	7800
ttggctgtgg tataataaat tattcataat gatagtagga ggcttgtagt gttaagaat	7860
agtttttgc tgaactttcta tagtgaatag agttaggcag ggatattcac cattatcgtt	7920
tcagaccac ctcccaaccc cgaggggacc cgacaggccc gaaggaatag aagaagaagg	7980
tggagagaga gacagagaca gatccattcg attagtgaac ggatccttgg cacttatctg	8040
ggacgatctg cggagcctgt gcctcttcag ctaccaccgc ttgagagact tactcttgat	8100
tgtaacgagg atttggaac ttctgggacg cagggggtgg gaagccctca aatattggtg	8160
gaatctccta cagtattgga gtcaggaaat aaagaatagt gctgttagct tgcctaatgc	8220
cacagccata gcagtagctg aggggacaga tagggttata gaagtagtac aaggagcttg	8280
tagagctatt cgccacatac ctagaagaat aagacagggc ttggaaggga ttttctata	8340
agatgggtgg caagtgttca aaaagtagtg tgattggatg gcctactgta agggaaagaa	8400
tgagacagc tgagccagca gcagataggg tgggagcagc atctcgagac ctggaaaaac	8460
atggagcaat cacaagtagc aatacagcag ctaccaatgc tgcttgtgcc tggctagaag	8520
cacaagagga ggaggagggt ggttttccag tcacacctca ggtaccttta agaccaatga	8580
cttacaaggc agctgtagat cttagccact ttttaaaaga aaagggggga ctggaagggc	8640
taattcactc ccaaagaaga caagatatcc ttgatctgtg gatctaccac acacaaggct	8700
acttccctga ttagcagaac tacacaccag gccagggggt cagatatcca ctgaccttg	8760
gatgggtgcta caagctagta ccagttgagc cagataagat agaagaggcc aataaaggag	8820
agaacaccag cttgttacac cctgtgagcc tgcattgggt ggatgaccgg gagagagaag	8880
tgtagtagtg gaggtttgac agccgcctag catttcacga cgtggccga gagctgcatc	8940
cggagtactt caagaactgc tgacatcgag cttgctacaa gggactttcc gctggggact	9000
ttccaggag gcgtggcctg ggcgggactg gggagtggcg agccctcaga tcctgcatat	9060
aagcagctgc tttttgcctg tactgggtct ctctggttag accagatctg agcctgggag	9120

ctctctgggt aactaggga cccactgctt aagcctcaat aaagcttgcc ttgagtgtt 9180
c 9181

<210> 15
<211> 29
<212> RNA
<213> Homo sapiens

<400> 15
ggcagaucug agccugggag cucucugcc 29

<210> 16
<211> 52
<212> RNA
<213> Homo sapiens

<400> 16
uuuuuuaggg aagaucuggc cuuccuacaa ggaaggcca ggaauuuuc uu 52

<210> 17
<211> 9413
<212> DNA
<213> Homo sapiens

<400> 17
ttggggcgga cactccacca tagatcactc ccctgtgagg aactactgtc ttcacgcaga 60
aagcgtctag ccatggcgtt agtatgagt ttgtgcagcc tccaggacc cccctccgg 120
gagagccata gtggtctgcg gaaccggtga gtacaccgga attgccagga cgaccgggtc 180
ctttcttggg tcaaccgcgt caatgcctgg agatttgggc gtgccccgc gagactgcta 240
gocgagtagt gttgggtcgc gaaaggcctt gtggtactgc ctgatagggt gcttgcgagt 300
gccccgggag gtctcgtaga cgtgcataca tgagacacaa tcttaaacct caaagaaaaa 360
ccaaacgtaa caccacccgc cgcccacagg acgttaagtt cccgggcggt ggtcagatcg 420
ttgttgaggt ttacctgttg ccgcgcaggg gccccaggtt ggggtgtgcg gcgactagga 480
agacttccga gcggtcgcaa cctcgtgga ggcgacaacc tatccccaag gtcgcccggc 540
ccgagggtag gacctgggct cagcccgggt acccttgccc cctctatggc aacgagggta 600

tggggtgggc aggaatggctc ctgtcacccc gtggtctctcg gcctagtgtg ggcacacacag	660
acccocggcg taggtgcgcgt aatttgggta aggtcatcga tacccttaca tgcggcttgc	720
ccgacotcat ggggtacatt ccgcttctcg gcgccccct agggggcgct gccagggccc	780
tggcacatgg tgtccgggtt ctggaggacg gcgtgaacta tgcaacaggg aatctgcccg	840
gttgcctctt ctctatcttc ctcttagctt tgcgtctctg tttagaccac ccagcttccg	900
cttacagggt gcgcaacgtg tccgggatat accatgtcac gaacgaactgc tccaactcaa	960
gtattgtgta tgaggcagcg gacatgatca tgcacacccc cgggtgcgtg cctctgcgtcc	1020
gggagagtaa tttctccgtg tgcgggtag cgcctacccc cagctcgcg gccaggaaaca	1080
gcagcatccc caccacgaca atacgacgcc acgtcgattt gctcgttggg gcggtcgtc	1140
tctgttccgc tatgtacgtt ggggatctct gcggatccgt tttctctctc tcccagctgt	1200
tcaccttctc acctcgccg tatgagacgg tacaagattg caattgtcca atctatcccg	1260
gccacgtatc aggtcacccg atggcttggg atatgatgat gaactggcca cctacaacgg	1320
ccctagtggc atcgcagcta ctccgcatcc cacaagccgt cgtggacatg gtggcggggg	1380
cccactgggg tgtcctagcg ggcttgcct actattccat ggtggggaac tgggctaagg	1440
tcttgattgt gatgtaact tttgctggcg ttgacgggca caccacgctg acagggggaa	1500
gggtagctc cagcaccocg agcctcgtgt cctggctctc acaaggccca tctcagaaaa	1560
tccaaactgt gaacaccaac ggcagctggc acatcaacag gaccgctctg aattgcaatg	1620
actccctcca aactgggttc attgctgcgc tgttctacgc acacaggctc aacgcgtccg	1680
ggtgcccaga gcgcattgct agctgcgcc ccctcgatga gttcgcctag ggtgggggtc	1740
ccctacacta tgatatgcct gagagctcgg accagaggcc atattgctgg cactacgcgc	1800
ctcagccgtg cgggatctg cctgcgtcgc aggtgtgtgg tccagtgtat tgcctcactc	1860
cgagccctgt tgtagtgggg acgacgcata gtttcggcgc tctactgtat agctgggggg	1920
agaatgagac agacgtgctg ctacttagca acacgcggcc gcctcaaggc aactgggttg	1980
ggtgcactg gatgaacacg actgggttca ccaagacgtg cggggggccct ccgtgcaaca	2040
tggggggggt cggcaaccaac accttgggtc gccccacgga ttgcttccgg aagcaccocg	2100
aggccaacta cacaaagtgt ggctcggggc cctgggtgac acccaggtgc atgtgtgact	2160
accatacag gctctggcac taacctgca ctgttaactt tacgtctttt aaggtaaggga	2220
tgtatgtggg gggcgtggag cacaggctca atgctgcatg caattggact cgaggagagc	2280
gctgtgactt ggggacacgg gataggtcag aactcagccc gctgctgctg tctacaacag	2340
agtggcagat actgocctgt tcttcaacca cctacccggc cctgtccact ggctgtatcc	2400
atcttcaacg gaacatcgtg gacgtgcaat acctgtacgg tatagggtcg gcagttgtct	2460
ccttgcgaat caaatgggag tatatcctgt tgcctttctc tcttctggcg gacgcgcgcg	2520

tctgtgctg cttgtggatg atgctgctga tagcccaggc tgaggccacc tttagagaacc 2580
 tgggtggctc caatgcggcg tctgtggcgg gagcgcatgg ccttctctcc ttctctgtgt 2640
 tcttctgcgc cgctgggtac atcaaaggca ggctggtccc tggggcggca tatgtctctt 2700
 atggcgtatg gcggttgctc ctgctcttgc tggccttacc accacgagct tatgccatgg 2760
 accgagagat ggcgtgcatg tgcggaggcg cgggttttgt aggtctggta ctcttgacct 2820
 tgtcaccata ctataagggt ttctctgcta ggctcatatg gtggttacaa tattttatca 2880
 ccagagccga ggcgcacttg caagtgtggg tccccctct caatgttcgg ggaggccggg 2940
 atgccatcat cctccttaca tgcgcggctc atccagagct aatctttgac atcaocaaac 3000
 tctgtctcgc catactcggt ccgctcatgg tgcctcaggc tggcataact agagtgcggt 3060
 actttgtacg cgctcagggg ctcatccgtg catgcatgtt agtgcggaag gtgcgtggag 3120
 gccactatgt ccaaatggcc ttcatgaagc tggccggcgt gacaggtacg tacgtatatg 3180
 accatcttac tccactgcgg gattggggcc acgcgggcct acgagacctt gcggtggcag 3240
 tagagccogt cgtcttctct gacatggaga ctaaaactcat cactcggggg gcagacaccg 3300
 cggcggtgtg ggacatcctc tcgggtctac cagtctccgc ccgaaggggg aaggagatac 3360
 ttctaggacc ggccgatagt tttggagagc aggggtggcg gctccttgcg cctatcacgg 3420
 cctattccca acaaacgcgg ggcctgcttg gctgtatcat cactagcctc acaggctggg 3480
 acaagaacca ggtcgatggg gaggttcagg tgcctccac ccgaacgcaa tctttcctgg 3540
 cgaactcggt caatggcggt tgttggaagg tctaccatgg tgcggcctcg aagacctgg 3600
 ccggcccgaa gggccaatc acccaaatgt acaaccaatg agaccaggac ctgctcggct 3660
 ggccggcgcc ccccgggggc cgctccatga caccgtgcac ctggcgagc tcggaccttt 3720
 acttggtcac gaggcgatgt gatgtcgttc cgggtgcggc gcggggcgac agcaggggga 3780
 gcctgcttcc cccagggccc atctcctacc tgaagggtc ctccgggtgga ccaactgctt 3840
 gcccttcggg gcacgttgta ggcattctcc gggctgctgt gtgcacccgg ggggttgca 3900
 aggggttggg ctctcatacc gttgagtcta tggaaactac catgcggtct ccggtcttca 3960
 cagacaaetc atccctcccg gccgtaccgc aaacattcca agtggcacat ttacacgctc 4020
 ccactggcag cggcaagagc accaaagtgc cggctgcata tgcagcccaa gggtaacaag 4080
 tgcctgctct aaaccgctcc gttgcggcca cattgggctt tggagcgtat atgtccaagg 4140
 cacatggcat cgagcctaac atcagaactg gggtaaggac catcaccacg ggcggcccca 4200
 tcacgtactc cactctttgc aagttccttg ccgacggtgg atgctccggg ggcgcctatg 4260
 acatcataat atgtgatgaa tgccactcaa ctgactcgac taccatcttg ggcacggcga 4320
 cagtctctga tcaggcagag acggctggag cgcggctcgt cgtgctcgcc acgcgccg 4380
 ctccgggata gatcccggt ccacacccca acatcgagga agtgccctg tccaacactg 4440

gagagattcc cttctatggc aaagccatcc ccattgagcg catcaagggg ggaaggcacc 4500
 tcattctctg ccattccaag aagaagtgtg acgagctcgc cgcaaagctg acaggcctcg 4560
 gactcaatgc tgtagcgtat taccggggtc tcgatgtgtc cgtcataccg actagcggag 4620
 acgtcgttgt cgtggcaaca gacgctctaa tgacggggtt taccggcgac tttagctcag 4680
 tgatogactg caacacatgt gtcacccaga cagtogattt cagcttggat cccaccttca 4740
 ccattgagac gacaacgctg ccccaagacg cgggtgtcgc tgcgcagcgg cgaggtagga 4800
 ctggcagggg caggagtggc atctacaggt ttgtgactcc aggagaacgg cctcaggcca 4860
 tgttcgactc ctccgtcccg tgtgagtgtc atgacgcagg ctgcgcttgg tatgagctca 4920
 cgcccgctga gacctcgggt aggttgcggg cttacctaaa tacaccaggg ttgcccgtct 4980
 gccaggacca cctagagtto tgggagagcg tottccacagg cctcaccacc atagatgcc 5040
 acttcttctc ccagaccaaa caggcaggag acaacctccc ctacctggtg gcataccaag 5100
 ccacagtgtg cgccagggtc caggctccac ctccatcggt ggaccaaaat tggagtgctc 5160
 tcatacggct aaagcccaca ctgcattggc caacgcccct gctgtacagg ctaggagccg 5220
 ttcaaaatga ggtcaacttc acacacccca taaccaata catcatggca tgcattgtcg 5280
 ctgacctgga ggtcgtcaat agcacctggg tgcctagtagg cggagtcctt gcggtctcgg 5340
 ccgcgtactg cctgacgaca ggcagcgtgg tcatttgtgg caggatcctc ttgtccggga 5400
 ggcacgtgtt tattcccgac agggaaagtc totaccagga gttcgatgag atggaagagt 5460
 gtgcttcaca cctcccttac atcgagcaag gaatgcagct cgcgcagcaa ttcaaacaga 5520
 aggcgctcgg attgctgcaa acagccacca agcaagcgga ggctgctgct cccgtggtgg 5580
 agtcaagtg gcgagccctt gaggtcttct gggcgaaaca catgtggaac ttcatcagcg 5640
 ggatacagta cttggcaggc ctatccactc tgccctgaaa ccccgcgata gcatcattga 5700
 tggctttttc agcctctatc accagcccgc tcaccacca aaataccctc ctgtttaaca 5760
 tottgggggg atgggtggct gcccaactcg ctccccccag cgctgcttgg gcttctgtgg 5820
 gcgcggcgat tgccggtgcg gccgttgcaa gcataggtct cgggaaggta cttgtggaca 5880
 ttctggcggg ctatggggcg ggggtggctg gcgcactcgt ggcctttaag gtcattgagc 5940
 gcgagatgcc ctccactgag gatctggtta atttaactcc tgccatcctt tctcctggcg 6000
 cctggttgtt cggggtcgtg tgcgcagcaa tactgctgtg gcacgtgggc ccgggagagg 6060
 gggctgtgca gtggatgaac cggctgatag cgttgccttc gcggggtaac cagctctccc 6120
 ccacgcacta tgtgcccgag agcgacgcgg cggcgctgtt tactcagatc ctctccagcc 6180
 ttaccatcac tcagtgtcgt aagaggtctc atcagtggat taatgaggac tgctccacgc 6240
 cttgttccgg ctctgtgcta aaggatgttt gggactggat atgcacgggt ttgagtgaat 6300
 tcaagacttg gtcacagtc aagctctcgc cgcggttacc gggaatccct ttctgtcat 6360

gccaacgcgg gtacaaggga gtctggcggg gggatggcat catgcaaacc acctgcccatt 6420
 gtggagcaca gatcaccgga catgtcaaaa atggctccat gaggattgtt gggccaaaaa 6480
 cctgcagcaa cagctggcat ggaacattcc ccataaacgc atacaccacg ggcacctgca 6540
 cgcctccccc agcgcggaac tattccaggg cgctgtggcg ggtggctgct gaggagtacg 6600
 tggaggttac gggggtgggg gatttccact acgtgacggg catgaccact gacaacgtga 6660
 aatgcccatg ccagggttcca gccctgaat ttttcacgga ggtggatgga gtacggttgc 6720
 acaggatgac tccagtgtgc aaacctctcc tacgagagga ggtcgatttc caggtcgggc 6780
 tcaaccagta cctggctcggg tcacagctcc catgtgagcc cgaacccgat gtggcagatgc 6840
 tcaactccat gtcacacgac cctctcata ttacagcaga gacggccaag cgtaggctgg 6900
 ccagggggtc tccccctccc ttggccagct ctccagctag ccagttgtct gcgccttctt 6960
 tgaaggcgac atgtactacc catcatgact ccccgacgc tgacctcacc gaggccaacc 7020
 tctgtggcg cgaggagatg ggcgggaaca tcaccctgt ggagtcagaa aataaggtgg 7080
 taatctgga ctctttgatg ccgattcggg cggtggagga tgagagggaa atatacgtcc 7140
 cggcggagat cctgcgaaaa cccagggaagt tccccccagc gttgccata tgggcaacgc 7200
 cggattacaa cctccactg ctgagtgctt ggaaggaccc ggactacgtc ccccgggtgg 7260
 tacacgggtg ccttttgcca tctaccaagg ccccccaat accaactcca cggaggaaga 7320
 ggaaggttgt cctgacagag tccaccgtgt cttctgcctt gcggagagtc gctactaaga 7380
 cctttggcag ctccgggtcg toggccgttg acagcggcac ggcgaactgga cctcccgatc 7440
 aggcctccga cgcagcgcga aaaggatccg acgttgagtc gtactcctcc atgccccccc 7500
 tcgagggaga gccaggggac ccgacccca gcgacgggtc ttggtctacc gtgacggggg 7560
 aagctggtga ggaagtcgtc tgcgtctcaa tgtctctatac atggacaggt gccctgatca 7620
 cgcacatgcg tcggaggagag agcaagttgc ccatcaatcc gttgagcaac tctttgctgc 7680
 gtcaccacag tatggtctac tcacacaacat ctgcagcgc aagtctgcgg cagaagaagg 7740
 tcacctttga cagactgcaa gtccctggacg accactaccg ggaactgtctc aaggagatga 7800
 aggcgaaggc gtccacagtt aaggctaggg ttctatctat agaggaggcc tgcaaacctga 7860
 cgcctccaca ttccgccaaa tccaaatttg gctacggggc gaaggacgtc cggagcctat 7920
 ccagcagggc cgtcaaacac atccgctccg tgtgggagga cttgtgtgaa gacactgaaa 7980
 caccaattga taccaccatc atggcaaaaa atgaggtttt ctgcgtccaa ccagagaaag 8040
 gaggcgcgca gccagctcgc cttatcgtat tcccagacct gggggtacgt gtatgcgaga 8100
 agatggccct ttacgacgtg gtctccaccc ttccctcaggc cgtgatgggc cctcctacg 8160
 gattccagta ctctcctggg cagcgggtcg agttcctggt gaatacctgg aaataaaga 8220
 aatgccttat gggcttctca tatgacaccc gctgctttga ctcaacgggc actgagaatg 8280

```

acatccgtac tgaggaatca atttaccat gttgtgactt ggcccccgaa gccaggcagg 8340
ccataaggtc gctcacagag cggctttatg tcgggggtcc cctgactaat tcgaaggggc 8400
agaactgcgg ttatcgccgg tgccgcgcaa gtggcggtgt gacgactagc tgcggcaaca 8460
ccctcacatg ttacttgaag gccactgcgg cctgtcgagc tgcaaagctc caggactgca 8520
cgatgctcgt gaacggagac gaccttgtcg ttatctgtga gagtgcggga acccaggagg 8580
atgcggcggc cctacgagcc ttacgggagc ctatgactag gtattccgcc cccccgggg 8640
acccgcccc aaccagaatac gacttggagc tgataaogtc atgctcctcc aatgtgtcgg 8700
tcggcgaaga tgcataccggc aaaagggtgt actacotcac ccgtgacccc accacccccc 8760
tcgcacgggc tcgctggggag acagttagac acactccagt caactcctgg ctaggcaata 8820
tcacatgta tcgccccaco ctatggcgca ggatgattct gatgactcat ttcttctcta 8880
tccttctagc tcaggagcaa cttgaaaaag ccctggattg tcagatctac ggggcctgtt 8940
actccattga gccacttgac ctacotcaga tcattgaacg actccatggt cttagcgcac 9000
tttcaactca cagttactct ccagggtaga tcaatagggt ggcttcatgc ctacaggaaac 9060
ttgggggtacc gcctttgcga gtcctggagac atcgggcccag aagtgtccgc gctaagctac 9120
tgtccagggt gggggagggt gccacttgcg gcaagtaact cttaactcgg gcagtaaaga 9180
ccaagcttaa actcactcca atcccgctg cgtcccagct agacttgtec ggctggttgc 9240
ttgctggtta caacggggga gacatatatc acagcctgtc tcgtgcccga ccccgttggt 9300
tcattgtgtg ctaactccta ctttctgtag gggtaggcac ctacctgtcc cccaacgggt 9360
gaacggggag ctaaccacto caggccaata ggccattccc tttttttttt ttc 9413

```

<210> 18

<211> 328

<212> RNA

<213> Homo sapiens

<400> 18

```

uuggggggga cacuccacca uagaucacuc ccugugagg aacuauguc uucacgcaga 60
aagcgucua gcaaggcggu aguaugagug uugugcagcc uccaggaccc cccuccccgg 120
gagagccaua guggucugcg gaacggguga guacacggga auugccaggga cgaacggguc 180
cuuuuuugga ucaaccgcgu caaugccugg agauuuugggc gugcccccgc gagacugcua 240
gccgaguagu guugggucgc gaaaggccuu gugguacugc cugauagggg gcuugcgagu 300
gccccgggag guucuguaa cogugcau 328

```

<210> 19
 <211> 14
 <212> RNA
 <213> Homo sapiens

<400> 19
 auuugggugu gccc 14

<210> 20
 <211> 27
 <212> RNA
 <213> Homo sapiens

<400> 20
 gccgaguagu guugggucgc gaaaggc 27

<210> 21
 <211> 340
 <212> DNA
 <213> Homo sapiens

<400> 21
 atgggcggag ggaagctcat cagtggggcc acgagctgag tgcgtcctgt cactccactc 60
 ccatgtccct tgggaaggtc tgagactagg gccagaggcg gccctaacag ggctctccct 120
 gagcttcagg gaggtgagtt ccagagaaac ggggctccgc gcgaggtcag actgggcagg 180
 agatgccgtg gaccccgccc ttcggggagg ggcccggcgg atgcctcctt tgcgggagct 240
 tggaacagac tcacggccag cgaagtgagt tcaatggctg aggtgaggta ccccgaggcg 300
 gacctcataa cccaattcag accactctcc tccgccatt 340

<210> 22
 <211> 349
 <212> DNA
 <213> Homo sapiens

<400> 22

gaggaaagtc	cgggctcaca	cagtctgaga	tgattgtagt	gttcgtgctt	gatgaaacaa	60
taaatacaagg	cattaatttg	acggcaatga	aatatcctaa	gtctttcgat	atggatagag	120
taatttgaaa	gtgccacagt	gacgtagctt	ttatagaaat	ataaaagggtg	gaacgcggta	180
aacccctcga	gtgagcaatc	caaatttggt	aggagcactt	gtttaacgga	attcaacgta	240
taaacgagac	acacttcogcg	aaatgaagt	gtgtagacag	atggttatca	cctgagtacc	300
agtgtgacta	gtgcacgtga	tgagtacgat	ggaacagaa	gcggcttat		349

<210> 23

<211> 377

<212> DNA

<213> Homo sapiens

<400>	23	
gaagctgacc	agacagtcgc	cgcttcogtcg
cgcttcctctt	cgggggagac	ggcgaggagg
		60
gaggaaagtc	cgggctccat	agggcagggt
gccaggtaac	gcctgggggg	gaaaccacg
		120
accagtgcga	cagagagcaa	acgcgcgatg
gccgcgcaa	gcgggatcag	gtaagggtga
		180
aagggtgcgg	taagagcgca	cgcgcggcgt
ggtaacagtc	cgtggcacgg	taaactccac
		240
ccggagcaag	gccaaatagg	ggttcataag
gtacggcccg	tactgaacc	gggtaggctg
		300
cttgagccag	tgagcgattg	ctggcctaga
tgaatgactg	tccacgacag	aaccggcctt
		360
atcggtcagt	ttcacct	
		377

<210> 24

<211> 38110

<212> DNA

<213> Homo sapiens

<400>	24	
ccaccggtta	cgatcttgcc	gaccatggcc
ccacaatagg	gcgggggaga	cccgcggtca
		60
gtggtgggog	gcacggtcag	taacgtctgc
gcaacacggg	gttgactgac	gggcaatata
		120
ggctccatag	cgctggccgc	ggatacagta
aaggagcatt	ctgtgaocga	aaagacgcc
		180
gacgacgtct	tcaaaacttc	caaggacgag
aaggtcgaat	atgtcgacgt	cgggttctgt
		240
gacctgcctg	gcacatgcga	gcacttcacg
attcggcctt	cgccctttga	caagagcgtg
		300
tttgacgacg	gcttgccctt	tgacggctcg
tcgatttcgg	ggttccagtc	gatccacgaa
		360
tcgcacatgt	tgcttcttcc	cgatcccgag
acggcgcgca	tcgacccggt	cgcgcggggc
		420

aagacgctga atatacaactt ctttgtgcac gaccogttca ccttgagacc gtactccgcg	480
gaccgcgcga acatcgcccg caaggccgag aactacctga tcagcactgg catcgccgac	540
accgcatact tcggcgccga ggccgagttc tacattttcg attcggtgag cttcgactcg	600
cggcgcaacg gctcctttta cgaggtggac gccatctcgg ggtggtggaa caccggcgcg	660
gcgaccgagg ccgacggcag tcccaaccgg ggctacaagg tccgccacaa gggcgggtat	720
ttccagtggt cccccaacga ccaatacgtc gacctggcg acaagatgct gaccaacctg	780
ataaactcgg gottcatcct ggagaaggcg caccacgagg tgggcagcgg cggacaggcc	840
gagatcaact accagttcaa ttogctgctg cagccgcgg acgacatgca gttgtacaag	900
tacatcatca agaacaccgc ctggcagaac ggcaaaacgg tcaogttcat gcccaagcgg	960
ctgttcggcg acaacgggtc cggaatgcac tgtcatcagt cgctgtggaa ggaaggggcc	1020
ccgctgatgt acgacgagac gggttatgcc ggtctgtcgg acacggcccg tcattacatc	1080
ggcgccctgt tacaccacgc gccgtcgctg ctggccttca ccaaccgcac ggtgaactcc	1140
tacaagcggc tggttcccg ttaacgagcc ccgatcaacc tggttctatag ccagcgcaac	1200
cggtcggcat gcgtgcgcac ccgatcacc ggcagcaacc cgaaggccaa gcggctggag	1260
ttccgaagcc ccgactcgtc gggcaaccgg tatctggcgt tctcgcccat gctgatggca	1320
ggcctggagc gtatcaagaa caagatcgag ccgcaggcgc ccgtcgacaa ggtatcttac	1380
gagctgccgc cggaaaggcg cgcgagtac ccgcagactc cgacccagct gtcagatgtg	1440
atcgaccgtc tcgaggcgga ccacgaatac ctacccgaag gaggggtgtt cacaaacgac	1500
ctgatcgaga cgtggatcag tttaacgcgc gaaaacgaga tcgagccggt caacatccgg	1560
ccgcatccct acgaattcgc gctgtactac gaogtttaag gactcttcgc agtccgggtg	1620
tagaggggag cggcgtgctg tgccagggcg ggcgtcgagg tttttcgatg ggtgacggtg	1680
gccggcaacg gcgcgccgac caccgctcgc aagagcccg ttaagaacgt tcaaggacgt	1740
ttcagccggg tgccacaacc cgcttggtcaa tcatctcccg accgccgagc gggttgtctt	1800
tcaatgcgcg cgaactcaaa gccacgtcgt cgcacaggcg tgcgtcgcg gccggttcag	1860
gttaagtgtc ggggattcgt cgtgcggcg ggcgtccacg ctgaccaacg ggcagtcga	1920
ctccgaaca ctttcgcac tacgccttt gccgcgcgcg taccctagat gtatgtgtcc	1980
aggaattccc caccgtcgtc gtttcgccag ccggccgcga ccgcgaccgc attgagctgg	2040
cgccccgggt ccggcagctg gtcggtggcg ttgcgcgcga ccaacaccag ccgcttgccg	2100
gcccggttg gctgcagcca ggccctgacg agcagctcca cgtcggtcgc ggggaaccga	2160
tggcgccgc cgtatgacac cagggaattgc agcgtcgagg tgtgtgcaag ggcgggaacc	2220
tgggtgcgat gctgtagctg cagcaactgc acggtccatt cgtatgtcgc cagtccgcgg	2280
cggcccaagt tgggtgtgtg gtlggggctg gcaccgcgcg gcaaccgcct ggactcgata	2340

cgggccttga tgcggcgaat ctgcgcacc gaggcagcgg acacaccgtc gggcggatac 2400
 cgggttttgt cgaccatcgg taggaatcgc tgaccacaact cggcatcgcc ggcaaccggc 2460
 tgtgcgcgta gcaggccctg gatctcccat ggctgtgccc actgctcgta gtagcgggcg 2520
 taggacccca gggtcgggac cagcgggaccg ttgcggccct cgggtcgcaa attggcgctg 2580
 agctccagcg gcggatcgac gclgggtgtc ccagcagcg ccggaaccgg ctccggcgatc 2640
 gatgtcgacc atttcaccgc ccgtgcctcg tcgacgcggg tggccggctc acagacgaac 2700
 atcagctcgg catcgacacc gtagcccaac tcggcaccac ccagccgacc catgcgcgatg 2760
 acccgcatgg ccgcggggcg gcgactgtcg tcgggaaggc tggccgggat catgaactcc 2820
 agcgcggcct gcgacaccgc caccacacc gacgtcaacg ccggcgacac ctccgtgacc 2880
 tcgagcaggc cgagcaggto ccggaaccg atgcgggcca gctctcgacg acgcagcgtg 2940
 cgcgcgcggc cgatggcccg ctccgggtcg gggtagcggc tcgccgaggc gatcagcgcc 3000
 cgagccacgg ccgggggctc ggtctcgagc agcttcgggc ccgacggccc gtctctgtac 3060
 tgtctgtaga ccgcggcgcg gcgcatcaac agatccggca catacgccga ggtacccaag 3120
 acatgcgatg gccgcttgcc caccgcggcg ttgtcccgca gcgtggccag gtaaccagctt 3180
 tcgggtggca gcgcctcact gagccgcgg taggccaagc gtccgcgcgc gggatcgggg 3240
 gcatacgaca tccagtcacg cagcctgggc agcagcacc actgcacccg tccgcgcggc 3300
 ccgctttgat tgaccaacgc cgacatgtgt ttcaacgcgg tctcggtgcc ctctgtagcc 3360
 agcgcggcca gccggcgccc cgcggcctcc aacgtcatgc cgtggcgcat tcccaaccgg 3420
 gtccggccga tcgattccag cagcggttga tagaagagtt tgggtgttaa ctccgacacc 3480
 cgcacgttct gcttcttgag ttctcccgcc agcaccocgg ccgcatcggt tcggccatcg 3540
 ggccggatgt gggcgcgcgc ccgacggcag cgcactgcct cctcgcttc gggatcgga 3600
 agcagggtggg tgcgcttgag ccgctgcaac tcagtcggt gctcgagcag cctgaggaa 3660
 tcatacgacg cggctatggt ccgcgcgtcc tcacgcccga tgtagccgcc ttgcaccaac 3720
 gccgccaatg cgtccaccgt ggaogccacc cgtaacgact cgtcgctacg ggcatagaac 3780
 agctgcagta gctgtacggc gaaotccacg tcgcgcaato ccgcgcgtcc gagtttgagc 3840
 tcggggccgc ggaatccggc gggcaccagc tgctccaccc gccgcgcgat ggcctgcacc 3900
 tcgaccacaa agtcttcggc ctgcgaggct ccgacacaca tcggcatcaa ggcggtcagg 3960
 taacgctcgc caagtccgcg gtcgccaacg actggccgtg ctttcagcaa ccctgaaac 4020
 tccaggtctc tggccacgag ctggtagtag gcgatgtgcg actcgagcgt acggaccagc 4080
 tcccgtttgc gccctccggc acgcaggcgcg gcgtccacct cgaataaggc ccgcgagggc 4140
 acccgcatca tctcgctggc cagcgcgcgg ttgcgcgggt ccgagcgctc ggcaaccgaat 4200
 atgacatcga cgtcgctgac gtagttcagt tcgcgcgca cgcacttgcc catcgcgatg 4260

accgccaggc gcggtggcgg gtgctgcgc caccgcctcg cctcgccacc gccgagcgcc 4320
 gccgccagag cggcgctcgc ggcgctcgc aggcgtgcgc ccaccacggg gaatggcagc 4380
 accggttcgt cctgcaccgt cggcgccagg tcgagagcgg ccagcattag cagctagtgc 4440
 cgggtactgg ttgcgaatgc gtgcacgagc gagcccgcca taccctcaga ttctcgcagc 4500
 cactgcagca acgaccgctg cagctgggtca tgggacggca gtgtgacctt gccccgcagc 4560
 aatttcacag actgcggatg ggcgaccagg tgatcgccca accccagcga cgagcccgagc 4620
 accgagaaca gccgcccgcg cagactgcgt tcgcgcagca gagccgcgtt gagctcgctc 4680
 catccggtgt ctggattctc cgacagccgg atcaaggcgc gcagcgcgcc atcgcgcttc 4740
 ggagcgctg acagcgacca cagcaggctg acgtgcgcct gatcctcgtg ccgatccacc 4800
 ccagctgag ccagacgctc accagcaggg gggcacaact atccgagccg gccaacgctg 4860
 ggcaacttc gccgctgcgt ggcgagtttg gtcaagacca cgacggttag ccaaagcgcg 4920
 tcggcgctcg atcaaccggt agatctgggc tacagcgaca ggtaggtgcg cagctcgtat 4980
 ggcgtagcgt ggctgcggta gttcgccacc tcctgctgct tgttcgcgaa gaaaagtca 5040
 aaaacgtgct cccccaaggc ctccgcgacg agttcggagg cctccatggc gccgagcgca 5100
 ctatccaaac tggcggcgaa ttctcggtac cccatcgctc gggttctc tcgggtgagc 5160
 tccatacgt tgctctgcgc ctgcgggccc agcagctaac ccttctctac accccgcaat 5220
 ccccgcgcca gcagcagcgc gaatgtcaga tagggattgc acccggaatc agggctcgct 5280
 acttcgaccc gccgcgacga ggtcttgctc ggcggtgaca tcggcacccg cactaggggc 5340
 gatcggttgg cggcccccca cgaagcgccc gtggggcgtt cgcgcgctg caccagccgc 5400
 ttgtaagagt tgacccactg atttgtgacc gcgctgatct cgcaagcgtg ctccaggatc 5460
 ccggcgatga acgatttacc cacttccgac agctgcagcg gatcatcagc gctgtggaac 5520
 gcgttgacat caccctcgaa caggctcatg tgggtgtgca tcgccgagcc cgggtgctgg 5580
 ccgaatggct tggcgcatga cgaagccggc gcgcctctt ccagcgcgac ttctttgatg 5640
 acgtagcgga aggtcatcac gttgtcagcc atcgacagag cgtcggcgaa ccgcaggtgc 5700
 atctcctgct ggcggggtgc gccttcgtga tggctgaact ccaccgagat gcccatgaat 5760
 tcacgggcat cgatcgctg cggcggaagg ttcaaggcgg agtcgtgcac cgcttggtgc 5820
 aaatagccgg cgttgctcac cgggacgggc accgaccgct cctcggttcc ggggttgagc 5880
 aggaagaact cgatttcggg atgcacgtag caggagaagc cgagttcgcc gcccttcgtc 5940
 agctgcgcc gccaacacgt ccgcgggttc gccacagacg gcgagccgct cggcatgggt 6000
 atgtcgcaaa acatccgcgc tgagtgggtg tggcggaac tggtgccca gggcagcacc 6060
 tggaaggtgc acgggtccgg gtgcgccacc gtatcggatt ccgagaccgc cgcaagcccc 6120
 tcgatcgagg atccgtcgaa gccgatgctt tctcgaagg cgcctcgag ttcggtcggg 6180

gcgatggcga cgcacttgag gaaacgcagc acgtctgtga accacagccg gacgaagcgg 6240
 atgtccggtt ctccaggggt acgaagaacg aatcccttct gtccgtccat acctcgaaac 6300
 gtatgcactg tctgttaaaa ccgtgttacc gatgcccgcc cagaagcggt gccggcgccg 6360
 ccgcgaaggg agtcgcgcgt gagtccaggg ccgcgcaccc agactcgtcg gccgcaaggt 6420
 ccgcgcgaga aaatagtgcg tcaccgcaga gtccacacac tggttgccat cgaacacccg 6480
 agtgtgttgg gtgcgcgtcg aggtgatcag ccgtgcgcgc agctggcggg ccaggtctac 6540
 ccgcgactga tacggagtggt ccgggtcgtg ggtggtggac accacgacga ccttgccacg 6600
 ccgcggccgg gccgcgggggt gccggcgtcg cgttgccggc accggccaca gccgcacag 6660
 atccgcgggg gccgatccgg tgaactgccc gtactgaagg aacggggcga cctgacggat 6720
 ccgttggtcg gccgccaacc agcccgctgg atccggccgt gtggcgccat ccagccacgg 6780
 gaccgcgttg aacgcgtcct ggtcgttgct gtagtgcccg tctgcatccc gccgctcata 6840
 gtccgtcgga agcaccagca agtcgcgcgc gtccgtcccg ccgtccagcc ccagcagacc 6900
 actggtcagg tacttccagc gctgagggct gtacagccgc ttgatggtgc ccgtcgtcgc 6960
 gtccgcgtag ctacggccac gtggatccga cgtcttacc gccctctgca ccagcgggtc 7020
 aaccagggcg tggtagcgggt tgaccactg gccgcagtcg gtgccacag gccgcggcgg 7080
 ccagcggggc cagtcggcgg ccgtagtcatt gaaagcggtc tgaatcccg ccatttgggt 7140
 gatgctttcc tcgattgggc taacggcgtg atccgatagc ccgtccagga ccacgcgccg 7200
 cccatgagta ccgaacgcgt ccaggttaag ggtgcccaac tccgtgcctg agctgtatcc 7260
 gaggtagttg atctgatcgt caccctaacc ttggcgaacc atgtccatgt ccctgcgcac 7320
 ggacgcggta ccgatattgg ccaagaagct gaagcccatc ccgtcaacac agtccgtggc 7380
 caactgcgg tagacctgtt ccagcgtgggt gaccccgcc ggactgtagt ccggccatcg 7440
 atccgcgcgg tacgcgtcg actccgcgtc ggtgcgacac ccgaacgcag gggctcagtg 7500
 gccgacccct ctccgggtcg agcccaacc gtccgaagtg ccgagaatgt ccgtgctggc 7560
 gatccgcgggt gccatagcgg ccacatgtgc gaccgcgcac gccccgggtc cccacggatt 7620
 gaccagcagt gtcctgaatc gctgtccgt ccgcggggac ccgatacccg ccaacttcgc 7680
 ttgtgtccca ccgggttggt ccgtatgcac ggggacggac accgtccgcg agcgtgcagt 7740
 ggaatttcgc ctggtgtcgg ccgatgaact gccgcagctg ttccaaactc gtgcggcgcc 7800
 ccgcacccgc gccacccggg ttgtgccggc gccgggttct tcagtccgcg ccgccaaccg 7860
 gggcgctgct aggggcagtc ccgcgagcag caaccgaag gacagcagc ccagctcaa 7920
 ccgtctcggc gccacatggt ccgccatcgt ctaccggcg aatacctgtg accgcgcgaa 7980
 atgatccac ctctgtttct tcgccccgct agcacttgcc gccgtggggc ggcgtggtgc 8040
 ccgcgattaa ataccgcgtc acgtactcgt caatgcagct gtccgccctg aataccaccg 8100

tgtgctgggt tccgtogaag gtcagcaacg aaccgcgaag ctggttcgcc aggtcgaccc 8160
 cggccttgta cggcgctgcc gggcatggg tggtagatac caccacgctc ggcacatggc 8220
 cgggcgcga gaaggcatgg ggcctgactg tgggtggcac cggccagaac gcgcaggtgc 8280
 ccaggcgccg atcaccggtg aactcccggt agctcatgaa cggctcgatc tcccgggcgc 8340
 ggcggtcttc gtogatgacc ttgtcgcat cggtaaccgg gggctgatcg acgcaattga 8400
 tcgccaccgg cgctgcaccg gaattgttgt agcgccgctg cgagtcocga cgcattgaca 8460
 tgtcgccag agccagcagg gtgtctccgc gattgtcgac cagctccgac agcccgctcg 8520
 tcaagtgtt ccacagatcc ggtgagtaca gcgccataat ggtgccacg atggcgctgc 8580
 tataactcag ccgcgcgga tccttcgtgc gcgcggcct gctgatctc gggttgtccg 8640
 ggtcgaccaa cggatcgacc aggtgttgtt agacctcgac ggctttggcc ggtcgccgc 8700
 ccaggcgga gcccgcgctc ttggcgagct cggcggcata gttgttgaa cgctcctgga 8760
 agcccttgcc ctgcgcgacg tccgcctcga tgggatcggc attggggtcg acggcaccgt 8820
 cgagaatcat tgcccgcacc cgctgcgga attcctcggc atacgcggag cgcattccgg 8880
 tgccgtacga gtgacccagg taggtcagct tgtctcgcc caacgcgcgg cgaatggcat 8940
 ccaggctcct ggcgagctt accgtccga catgggccc aaagtcttg ccatcttgt 9000
 ccacacagc accgacgaat tgcttgggtc cgtctcgat gtgcgccaca cctcccgcc 9060
 tgtagtcaac ctgcggtcgc gccgcagcc ggtcgttgtc ggcattggag ttgcaccaga 9120
 tcgcggcgc ggaagcgcgc accccgcggg ggtogaaccc aaccaggctg aacctctgt 9180
 gcaccgctt cggcaatgtc tggaagacgc ccaaggcgcc ctcgatacgc gattcgccg 9240
 gtccaccggg atttatgacc agcgaaccga tcttgtctcc cgtcgccga aagcgaatca 9300
 gcgccagcgc ccgacgtca ccacggggc ggtcgtatgc gaccgtgaca gcgagcttg 9360
 cgcataacgc gcccgccggg atctttactt cggggttga cgaccggcac ggtgtccact 9420
 ccaccggctg gccagcttc ggcctcccca tacgagcgcg tccccgacc acgggatgc 9480
 agcccacaag aaccaagcc accggcgga gcgcggccca gatcaacagc atgcgcgga 9540
 tcttgtcgcg gcgagacagc ctcatgccca caatgtgcg agagcagacc cgagatcctg 9600
 gccagcgcc accgtggcc gactaacggc ccgtgcccag cagtctgcc atcgccgatg 9660
 gcgaactcgt cggccatccc ccatacgtcc ggtaacagat ccgggcaaga caccgacccg 9720
 tcgaccggat ccggcagcgg cgctcgccgc tcggcggtgc acaactgcga catcagggtt 9780
 gcgtggcac ccgctccacg ccggcatggt gcaccttgcc catcgccga gggcgatccc 9840
 cgatgcgctc cacccttcg acgaacccat ctccacggc ggtcgccgc agcgacgga 9900
 tgtggccga gatctccgag agttcggccc gcccgcccg gcagcgcaac ccgatgcgt 9960
 gcaagtgcag atcgatgtga ggttcaaggt tcagcgact gctggcaagc tttttccgaa 10020

accgcggcct cgccttgatc tggagtcaga acgcgtcacg cagccgggtca aaggcgtaac 10080
 ccattgctcga gcaaacatgc atggggtgag tggacgttcc cagacacagc aactggcgctc 10140
 caggccactg agcccgctgca tgcgcgatgg tatgcgatgg ggggccccgg gcgcgtctga 10200
 ggggaagaag tggcagactg tcagggtccg acgaaccogg ggaccctaac gggccaacag 10260
 gatcgaccgg accaccatta gggacagtga tgtctgagca gactatctat ggggccaata 10320
 ccccgaggag ctccggggcg cggaccaaga tccgcaccca ccacctacag agatggaagg 10380
 ccagcggcca caagtgggccc atgctgacgg cctacgacta ttgcacggcc cggatctctg 10440
 acgagggcgg catcccggtg ctgctgggtg gtgattcgcc ggccaacgtc gtgtacggct 10500
 acgacaccac cgtgcggatc tccatcgacg agctgatccc gctggtccgt ggctgtgtgc 10560
 ggggtgcccc gcacgcactg gtcgtgcgag acctgcggtt cggcagctac gaggcggggg 10620
 ccacgcgcgc gttggccgccc gccaccgggt tectcaagga cggcgcgcca catgcggtca 10680
 agctcgaggg cggctgagcgg gtggccgagc aaatgcgctg tctgaaccgg cggggcatcc 10740
 cgggtgatgg acacatcgcc ttaccocgc aaagcgtcaa cacttgggc ggcctccggg 10800
 tgcagggcgg cggcgacgcc gccgaacaaa ccatcgccga cgggatcgcc gtgcgcgaag 10860
 cgggagcgtt tgcgctgtg atggagatgg tgcgcgcga gttggccacc cagatcaccg 10920
 gcaagcttac cattccgacg gtcgggatg gcgctggggc caactgcgac ggcaggtcc 10980
 tggtatggca ggacatggcc ggggttcagc ggcgaagac cgcgcgcttc gtcaaacggt 11040
 atgcgatgt cgggtgtgaa ctacgcgctg ctgcaatgca atacgcccga gaggtggcgg 11100
 ggggggtatt ccccgctgac gaacacagtt tctgaccaag cgaatcagc ccgatgcgcg 11160
 ggcatctcgg tggcgccctg gatgcgctg acgcgggatt gccgcgcggg acgcgccagc 11220
 gggaccatc ggctgcgctg tcgcgggttg agcccggggt gagccacagc attcgatgtg 11280
 cccaacacca tccgccacag cccaattgat gtggcactct atgcatgcct atcccagacc 11340
 aaccaccacc ggcgcgacgc atcatgacgg gaggcgaaga tgcagtaga ggcgccaga 11400
 ccagcgccgc atctggaggt cagcgcgaag ttgcagctga tgcagtcgac ggtgtcgccg 11460
 tggttcgagg gcacgcgcgc ggtggttcgc gtgcagcagt cgcgcaccca gcagctcgac 11520
 gcggtgtact tcgacacacc gtgcacgac ctggcgcgca accagatcac ctgcgcgcg 11580
 cgcacccggg gcgcgcagc cggctggcat ctgaagctgc cggccggacc cgaacaagcg 11640
 accgagatgc gagcaccgct gtcgcgcatc ggcgacgctg tgcggcgga gttgttgatg 11700
 gtggtgtcgg cgatcgtccg cgaccacgg gttcagcgg tcgcgcggat cagcactcac 11760
 cgcgaagacc agatcctgta cggcgccggg ggcgcgcgc tggcggaatt ctgcacgac 11820
 gacgtccacc catggtcgcc cggggcattc cagcgcgctg gtgcagcgg caacggccct 11880
 gccgaacagc agtggcgga atgggaactg gaactggtca ccacggtgg gaccgccgat 11940

accaagctac tggaccggct agccaaccgg ctgctcgatg ccggtgcccgc acctgcccgc 12000
 caccggctcca aactggcgcg ggtgctcggt gogacctctc ccggtgagct gcccaaccgc 12060
 ccgcagccgc cggcggatcc agtacaccgc gcggtgtccg agcaagtoga gcagctgctg 12120
 ctgtgggacg gggccgtgcg ggcgcagccc tatgacgcgg tgcaccagat gcgagtgaag 12180
 acccgcaaga tccgcagctt gctgacggat tcccaggagt cgtttggcct gaaggaaagt 12240
 gcgtgggtca tcgatgaact gcgtgagctg gccgatgtcc tgggcgtagc ccgggacgcc 12300
 gaggtactcg gtgaccgcta ccagcgcgaa ctggaccgcg tggcgccgga gctggtacgc 12360
 ggccgggtgc gcgagcgctt ggtagaccgg gcgcggcgcg gataccagac cgggctcgcg 12420
 cgataactga tcgcattgcg gtcgcagcgg tacttccgtc tgctcgacgc tctagacgcg 12480
 cttgtgtccg aacgcgccca tgccacttct ggggaggaat cggcaccggg aaccatcgat 12540
 gcggcctacc ggcgagtcgg caaagccgca aaagccgcaa agaccgcggc gcaccaggcg 12600
 ggcgaccacc accgcgacga ggcattgcac ctgatccgca agcgcgcgaa gcgattacgc 12660
 tacaccgcgg cggctactgg ggcggacaat gtgtcacaag aagcgaaggt catccagaag 12720
 ttgttagggc atcatcaaga cagcgtggtc agccgggaac atctgatcca gcaggcata 12780
 gcccggaaca ccgcggcgga ggacaccttc acctacggtc tgcctacca acaggaagcc 12840
 gacttgccg agcgtgtccg ggagcagctt gaagccgcgc tgcgcaaact cgacaaggcg 12900
 gtccgcaaag caccggattg agcccgccag gggcggaaga gttggcctgt aagccggatt 12960
 ctgttccgcg ccgcaccagc caagctaaag gcggcacgga ggcgaccatc catctggaca 13020
 caccgttacc ggttgctcgc agcggcctac ccgcaggctc gggcgagcaa cctcgaagcg 13080
 cctgcgcggc cgcactttcg gtgcggcctt cttggccttg cttcgggtgg ggtttgccta 13140
 gccaccccg gcacccgaa tgctggtgcg ctcttacgc accgtttcac cttgccacc 13200
 acgaggatgg cggctctgtt tctgtggcac ttcccgca gtcacctcg attgcggtta 13260
 gcaatcacc tgctctgtga agtcggact ttctcgact cgaacctgaa cctogtgaat 13320
 ccacacaagc cctacgcgag ccgcggccgc ccagccaaat catccgcgac gaaccageta 13380
 ccccgctggg cgggtgtcgc gccagtgta ccgctggacg acacggttag toggacagcc 13440
 gatccggcg gcagtcctta tcgtggactg gtgacacggg gggacaaagc cgtogactcc 13500
 ggcgactggg accgcactgc tgcgaggtc agcgagtacg ttggcgact gctacctcgg 13560
 ctgatcacc ccggcgaggc gcgccggctg cgcaagctgt accgcgacga cggcctgttt 13620
 cgtctcagcg tcgatatggc atccaagcgg tacggcgccg ggcagtatcg atatttccat 13680
 gccccctatc ccgagtgtac gagcgtctca agcaggcgct gtatcccaaa ctgtgcgca 13740
 tagcgcgcaa ctggtgggcc aaactgggcc gggaggcgcc ctggccagac agccttgatg 13800
 actggttggc gagctgtcat gccgcgggcc aaaccgcatc cacagcgtg atgttgaagt 13860

acggcaccaa cgactggaac gccctacacc aggatctcta cggcgagttg gtgtttccgc 13920
 tgcaggtggt gatcaacctg agcgatccgg aaaccgacta caccggcggc gaggttcctgc 13980
 ttgtogaaca gcggcctcgc gcccaatccc ggggtaccgc aatgcaactt cgcgagggac 14040
 atggttatgt gttcacgacc cgtgatcggc cggtgcggac tagccgtggc tggctggcat 14100
 ctccagtgcg ccatgggctt togactattc gttccggcga acgctatgcc atggggctga 14160
 tctttcacga cgcagcctga ttgcacgcca tctatagata gcctgtctga ttcaaccaac 14220
 gcaccgacga tgccccatcg gcgtagaact cggcgatgct cagcgatgcc agatcaagat 14280
 gcaaccgata taggagcccc gaccggcat ccaacgccag ccgcaacaac attttgatgc 14340
 gcgtgacatg tgacaccacc agcacctcgc cgccttcgta gccaacgatg atccgatcac 14400
 gtcccgccgc aaccgcgcgc agcacgtcgt cgaagcttcc ccaaccggg ggcgtgatgc 14460
 tgggtgtcctg cagccagcga cgggtgcagct cgggatcgcg ttctgcggcc tccgcgaacg 14520
 tcagccctc cccggcgccg aagtcgggtct cgaccaggtc gtcacgacg accacgtcca 14580
 gggccagggc tctggcgcg gtcaccgcgg tctgtaagc ccgctgtacg ggcgaggaga 14640
 ccaccgcagc gatcccgccg cgcgcgcgca gataccggc cgcgcgacca acctggcgcc 14700
 accccacctc gttcaacccc gggttgcgcg gccccgaata gcggcggttc tccgacagct 14760
 ccgtctgcgc glggcgcaac aaaagtagtc ggggtgggtgt accgcggcg ccggtccagc 14820
 cgggagatgt cgggtgactgc gtcgcaacga ttttggcagg atccgcaccc gccgcagccg 14880
 attgcgcggc ggcgctccat cgtcattgg ccaaccggtc tgcatacgtg ttcggggcac 14940
 gcggaaccca ctctagattg atcctgcgaa actgggacgc caacgcctga gcctggacat 15000
 agagcttcag cagatccggg tgcttgacct tccaccgcc ggacatctgc tccaccacca 15060
 gcttgaggte catcagcacc cgggcctcgg tggcacctag ttccacggcg tcgtccaaac 15120
 cggctatcag gccgcggtat tcggcgacgt tgttcgtcgc ccggccgacg gcctgcttgg 15180
 actcggccag caccggtggag tgatcggcgg tccacaccac cgcgcgctat ccggccggtc 15240
 cgggattgcc ccgcatccgc ccgtcggctt cgatgacaa tttoactcct caaatccttc 15300
 gagccgcaac aagatogctc cgcattccgg gcagcgacac acttcactct cggcgcgccg 15360
 cgagatctgg gccagctcgc cggggccgat ctcgatccgg caggcaccac atcgatgacc 15420
 ttgcaaccgc ccggccctgc gcccgctccc ggcccgctgt ctctcgtaga gccccgcaag 15480
 ctccggatca agtgtccgc tcagcatgtc gcgttgcatg gaatgttggt gccgggcttg 15540
 gtcatattcg gcaagtgcct cgtccaaagc ctgctggcg cgcgccaggt ccggcccgaa 15600
 cgcttgagc gcccgcgact cggcggtctg ttgagcctgc agctcctcgc ggcgttccag 15660
 cactccagc agggcatctt ccaactggc ttgacggcgt tgcaagctgt cgagctcgtg 15720
 ctgcagatca gccaatgtct tggcgctcgt tgcaccgaa gtgagcaacg accggtcccg 15780

gtgcacacgc ttacgcaccg catcgatctc cgactcaaaa cgcgacacct ggcgcgtocaa 15840
 gtctctcgcc gogattcgca gggccgcgat cctgtcgttg gcggcggttg gctcgccctg 15900
 cacctgctgg taagccgccg gctgcggcag atgggtagcc cgatgcgcga tccgggtcag 15960
 ctccagatcc agcttcgcca attccagtag cgaccgttgc tgtgccactc ogggtttcat 16020
 gcctgatctc tcccagtttc gtgatcgagg ttccacgggt cggtgcagat ggtgcacaca 16080
 cgcaccggca gcgacgcgcc gaaatgagac cgcaacactt cggcggcctg gccgcacacc 16140
 gggaaattcg ttgccaatg cgcgacgtcg atcagggccca cttgcgaagc tcggcaatgc 16200
 tcgtggctg gatgatgtcg cagatcgccc gtaacgtacg cttgcacgtc cgcggcgccc 16260
 acggtggcaa gcaacgagtc cccggcgccg ccgcagaccg cgaccgcga caccagcagg 16320
 tcgggatccc cggcggcgcg cacaccggtc gcagtcggcg gcaacgcggc ctccagacgg 16380
 gcaacaaagg tgcgcagcgg ttccgggttt ggagtcctgc caatccggcc taaccgcgtg 16440
 ccgaccggcg gtggtaccag cgcgaagatg tcgaatcccg gctcctcgta aggggtcgcg 16500
 gcgcgcctcg ccgccaacac ctgcggcgcg gctcgtcgcg gtgcagacgc ctgaccccg 16560
 tctcggccca ccgcttcgac ggtaccgacg ctgcctatgg cgggcgacgc ccgctcgtgc 16620
 gccaggaact gcccggtacc cgcgacactc cagctgcagt gcgagtagtc gccgatatgg 16680
 ccggcacccg cctcaaagac cgtgcgccg accgcctctg agttctcgcg cggcacatag 16740
 atgacccact tctcgagatc ggccgctccg ggcaacgggt cgagaacggc gtgcagggtc 16800
 agaccaacag cgtgtccagc cgcgtcgacc acaccggcg accgcgagtc ggcgttggtg 16860
 tgcgcggtaa acaacgagcg accggtcccg atcaggcggt gcaccagcac accctttggc 16920
 gtgttgccg cgaccgtatc gacccacgc agtaacaacg ggtggtgcac caatagcagt 16980
 ccggcctggg gaacctggtc caccaccgcc ggctgcgt ccaccgcaac ggtcacccaa 17040
 tccaccacgt cgtcggggtc gccgcacacc agaccaccg aatcccacga ctgggcaagc 17100
 cgcggcggtt aggcctggtc cagcagctcg atgacatcg ccagccgcac actcatcgcc 17160
 gtctccacg ctttgccacc tcggcgatcg ccgccaccg caccggccac tccgggcgca 17220
 ccgcgcgccg caggtaccgc cgcgccaggc cgaagagggt gtcaccgcgg ccgaccgcga 17280
 ttctttgtct ctgcaaatag ttctgaatc cgtcagcatc ggcatgttg aacagtagca 17340
 aaggggcgcc accatcgacc acctcggcac ccaccgatct cagtcggccc accatctccg 17400
 cgcgcagcgc cgtcaaccgc accgcacatc ctcgggcagg ggcaacgcc cggggggcgc 17460
 agcaagcagc gatggcgtc agttgcaalg ttcccaacg ccagtgcgt cgtgcacgg 17520
 tcaaccgagc cagcagctct ggcgagccga gcgcgtagcc caccgcgaat ccggccagcg 17580
 accacgtttt cgtcaagcta cggagaccca gcacatcggg cagcgagtca tcggccaacg 17640
 attcgggctc gccgggaacc caatcagcga acgcctcgtc gaccaccagg atgcgtcccg 17700

gccggcgtaa ctcgagcagc tgctcgcgga ggtgcagcac cgagggtggg ttggtcggat 17760
 taccacagac gacaaggctc gcgtcgtcag gcacgtgcgc ggtgtccagc acgaacggcg 17820
 gctttaggac aacatggtgc gccgtgattc cggcagcgct caaggctatg gccggctcgg 17880
 tgaacgcggg cacgacgatt gctgcccgca ccggacttag gttgtgcagc aatgcgaatc 17940
 cctccgcgcg ccgacgcgag gggagcaett cgtcacgggt tctgccatga cgttcagcga 18000
 ccgcgtcttg ccgccggtgc acatcgtcgg tgctcggata cggggccagc tcgggcagca 18060
 gcgcggcgag ctgcgcggac aaccattccg ggggcgggtc atggcggagc ttgacggcga 18120
 agtcacgac gccggggcggc acatcctgat caccgtggta gcgcgcgcgc gcaacggggc 18180
 tagtgtctag actcgccaca gcgtcaaaca gtatgtggcc ggtgtgcggg ccaagaatcc 18240
 agagcaccgc cgacgcggtt tctacgcggc gacaaccgcg acatcacagg cagctaacag 18300
 ggcgctggcg gtgatgatcg tcaggccaag cagctgtgcc tgggcgatga gcacacggtc 18360
 gaatggatgt cgatggtgat ccggaagctc tcgggtgcgc agtgtgtgcg tggccaactg 18420
 acagcggcga cgtgcccgag ccgcgcattc gatcgggcac gtaagaagcc gatggctcgg 18480
 gcggcgggag cttgccgagg cggtagttga tcgcgatctc ccaggcaactg gcggccgaca 18540
 agagaatgct gttgcggagc tctgaacaa tcgcccggtt tctgttgacg gcattccgag 18600
 ccaaactggt gttgcgatga ggtagcgtt caccgggtga agcgttcagc cagctcgtct 18660
 gacaacggag cgtccaatc gtcgggcacg cggtaacgc catgtcaat gcctaaccgc 18720
 cgagtctcat gaggatgcag ccgcacaagc ttgtctacc gctgcgcgc gcgggcaatc 18780
 tcaacctctg ccgcgcgtag acgagccgca gcagctcgga caggcgtgtc ttgcctcgt 18840
 gaacgcgcac ccgcttcgca ggcgccaga ctttcgcgtc gaccacctgc tcaccaaact 18900
 tcgcgatcat cgcctgatac cacagcgcca acgggtagcg gttgtccaa ccgcttcgtc 18960
 aacgacaatg ggatcgtgac cgacacgacc gcgagcggga ccaattgcc gcctcctcca 19020
 cgcgcgcgcg caccggcgcg atcgtcgcgc ggtgaatgc ccgcagctggt gatcttcgat 19080
 ctggaaggca cgtgaccca ctcggcgcg gcgaatcgt acagcttcgc acacgcgcgc 19140
 aaccacatcg gtgcccagc acccgaggc gaacctgcca ctacatcgt cggcccgccc 19200
 atgatgaga ccctgcgcgc catggggctc gggaatccg ccgaggaggc gatcgtagcc 19260
 tacggggcgc actacagcgc ccgcggttg gcgatgaac gcttgttcga cgggatcggg 19320
 ccgctgctgg ccgacctgc caccgcgggt gtcgggctgg ccgtgccac ctccaaggca 19380
 gagccgacgc caccgcgaat cctgcgccac ttcggaattg agcagcaact cgaggctatc 19440
 gcgggcgcga gcaccgatg ctcgcgagcg agcaaggctc acgtgctggc ccacgcgcgc 19500
 gcgcagctgc ggcgcctacc cgagcgggtg gtgatggtcg gcgaccgcag ccacgacgct 19560
 gacggggcgc ccgcgcagc catcgacacg gtggtggtcg gctggggcta cgggcgcgc 19620

gactttatcg acaagacctc caccacogtc gtgacgcatg ccgccacgat tgacgagctg 19680
 agggaggcgc taggtgtctg atccgctgca cgtcacatct gtttgtacgg gcaacatctg 19740
 ccggtogcca atggcccgaga agatgttcgc ccaacagctt cgccacogtg gcttgggtga 19800
 cgcggtgcga gtgaccagtg cgggcaccgg gaactggcat gtaggcagtt gcgcgcagca 19860
 gcgggcggcc ggggtgttcg gagccacggc ctaccctacc gaccacccgg ccgcacaagt 19920
 cggcaccgaa cacctggcgg cagacctgtt ggtggccttg gaccgcaacc acgctcggct 19980
 gttgcggcag ctocggctcg aagccgcccg ggtacggatg ctcggtctat tcgacccacg 20040
 ctccgggaacc catgcgctcg atgtcgagga tccctactat ggcgatcact ccgacttoga 20100
 ggaggctctc gcgcgtcatg aatccgccct gcccggcctg cagcactggg tcgacgaacg 20160
 tctcgcgcgg aacggaccga gttgatgcc cgcctagcgt tctgtctgcg gcccggtctg 20220
 ctggcggttg ccttggtcgt ggtcgcgttc acctacctgt gctttacggt gctcgcgcgg 20280
 tggcagctcg gcaagaatgc caaaacgtca cgagagaacc agcagatcag gtattccctc 20340
 gacaccccg cggttccgct gaaaaccctt ctaccacagc aggatctcgc ggcgcgggac 20400
 gcgcagtggc gcgggtgtac ggcaaccgga cagtaccttc cggacgtgca ggtgctggcc 20460
 cgactcgcgc tgggtggagg ggaccaggcg tttgaggtgt tggccccatt cgtggtcgac 20520
 ggcggaacca ccgtcctggt cgaccgtgga tacgtgcgpc ccaggtggg ctgcacgta 20580
 ccaccgatcc ccgcctgcc ggtgcagacg gtgaccatca ccgcgcggct gcgtgactcc 20640
 gaaccagagc tggcgggcaa agacccttc gtcagagacg gcttccagca ggtgtattcg 20700
 atcaataccg gacaggctgc cgcgctgacc ggagtccagc tggctgggtc ctatctgcag 20760
 ttgatogaag accaaaccgg cgggctcggc gtgctcggcg ttccgcatct agatcccggg 20820
 ccgttctcgt cctatggcat ccaatggatc tcgttcggca ttctggcacc gatcggttg 20880
 ggctatttcg ctaacgcgga gatccggggc cgcgcgcggg aaaaagcggg gtgcgccaca 20940
 ccggacaagc caatgacggt cgagcagaaa ctgcgtgacc gctacggccg ccggcggtaa 21000
 accaacatca cggccaatac cgcagccccc gcctggacca ccgcgcagac caccacggcg 21060
 cggcgagat cggccacctt gggcgaccgg ccgtcgccca aggtggggcg gatctgcaac 21120
 tcatggtggt accgggtggg cccaccacgc cgcacgtcaa gcgcccacgc aaacgccgc 21180
 tcgacgacac cggcggtggg gctgggatgg cgggcggcgt cgcgcgcgca ggcgccgtacc 21240
 gcaccgcggg gcgaccaccc gaccaccggc gcgcagatca ccaccagcac gcgcgtgcgc 21300
 cgtgcgcaa catagtgtgc ccagtcatcc aatcgtgctg cagcccaacc gaatcggaga 21360
 taacgcggcg agcggtagcc gatcatcgag tcagggtgtg tgatggcacg atatccagc 21420
 accgcaggca gcgcgctcga agccgcccac agcagcgcca ccacctgggc gtcggcggtg 21480
 ttttcggcca ccgaectcag cgcggcacgc gtcaggcccg ggcgcgccag ctgggcgggg 21540

tcacgccgcg acagcgacgg cagcagccgt cgcgcgcgct cgacatcgct cgcgtccaac 21600
 aggtccgata tctggcgccg ggtgcgcgcc agcgaagttc cgcgcagcgc tgcccaggtg 21660
 gcgctcgcgg tggccgccac gggccaggac ctgccgggta gccgtgcag tgccgcgcg 21720
 agcaagccca cgcgcgcgac cagcaggccg acgtgtaccg caccggcgac cgggcgctca 21780
 cggtaggtga tctgctccag cttggcgccg gcccgaccga acagggccac cggatgacct 21840
 cgtttggggg cgcgaacac gacgtcgagc aggcagccga tcagcagccg gaaggccctg 21900
 gtctgccagg tcatgcaaa cactccggca gcgtcgaca cgtggctac gctcagctat 21960
 ttatgacctc atacggcagc tatccacgat gaagcgccca gctaccggg ttgccgacct 22020
 gttgaaccg gcggcaatgt tgttcggcg agcgaatgtc atcatcgacg tggcagtgcc 22080
 ggggtgctgg tatggcgtgc tggaaagccc ggtggacagc ggcaacgtct acaagcatcc 22140
 gttcaagcgg gcccggaaca cgggcacctc cctggcggtg gcgaccatcg ggaaggaaac 22200
 cgaccgagcg ctgatccggg gtgcggtgga cgtcgcgcac cggcagggtc ggtcgacggc 22260
 ctcgagccca gtgtctata acgccttcga ccgaagtgt cagctgtggg tggcgcgctg 22320
 tetgtaccgc tactctgtg accagcacga gttctgtac ggccactcg aagatgcac 22380
 cgcgcagccc gtctaccaag acgcacaaag gttagggacc acgctgcagg tgccggaggg 22440
 gatgtggcg ccggaccggg tcgcgttcga cagttactgg aagcgctcgc ttgatgggct 22500
 gcagatcgac gcgcgggtgc gcgagcatct tcgcgggggt gcctcggtag cgtttctccc 22560
 gtggcggtg cgcgcgtgg ccgggcgctt caacctgtt gcgacgagc gattcttgcc 22620
 accggagttc cgcgcgatga tgcagctgga gtggtcacg gcccgagcgc gtcgcttcga 22680
 gtggttactt tcgctgtac ggttagccga ccggctgatt ccgcatcggg cctggatctt 22740
 cgtttaccag ctttactgt gggacatgcg gtttcgcgcc cgacacggcc gccgaatcgt 22800
 ctgatagagc ccggccgagt gtgagcctga cagcccgaca ccggcgcgct gtgtcgctc 22860
 gccaggttca cgtcggcga totagagccg ccgaaaacct acttcgggt tgccctccga 22920
 atcaactgct tgatctgctc gagcagctca cgcataatcg cgcgcacgc atccaccgcy 22980
 gcatacaggt cggccttggt cgcggcgagc tggtcgagc toattggcg caccggcggt 23040
 gctgtctgct gcgcgcgcgt gtcgcttga aaccaggtc gtcacccac gaaccagaca 23100
 ctgccatata cggcgccccc ccgacacga agcacagcta gccggtgggc gcggacggga 23160
 tcgaaccgccc gaccgcgtgt gtgtaaaacc agagctctac cgtgagcta cgcgcccatg 23220
 accgcccgag gctacacgcc ttgcggccaa gcacccaaaa ccttaggcgc taagcgccgc 23280
 cagagcgtcg gtcacagcc gctgatcgcg aacttcaccc ggtgcttca tctcggcgaa 23340
 ccgaatgac cctgaccgat cgaccacaaa ggtgccccg ttagcgatgc cggcctgctc 23400
 gttgaagacg ccgtaggcct gactgacccg gcgctgtggc cagaagtccg acaacagcgg 23460

aaacgtgaat ccgctctcgc tcgccagat cttgtgagt ggtggcgggc ccaccgaaat 23520
 cgctagccgc gcgctgtcgt cgttctcaaa ctgggcagg tgatcacga actggtccag 23580
 ctgcctctgc cagatgccgc tgaacgcaa cggaagaac accaacgca cgttctttgc 23640
 accccggtag ccgcgcagg tgacaagctg ctgattctgg tcgcgaacg tgaagtacgg 23700
 ggcggtggct ccgacgttca gcatcagcgc ttgccagccc gcgatttcgg ctgtaccaat 23760
 ctgctggcgc tccagttgcc cagattgacc gacgaggtcg gcatcagccc agctgtgggc 23820
 gccgcctcgc caatctcgc gggcaataca tggccgggct ggcggtctt ggcgctcacc 23880
 acccaaatca caccgtctc ggcgagcggc ccgctcgc ccatcagggt gtccaccaa 23940
 tcgcctcgc catcacgca ccacaacagg acgacatga tgacctcgc ggtgtcttca 24000
 tcgagcaact ctccccgca cgtctcttcg atggccgc ggatgtctg gtctgtctt 24060
 tcgtccacgc ccatctctg gataagttgg tctcgttga tgcccaatt gcggcgtag 24120
 ttcgaggcgt gatccgcgc gaccacgtg gaacctcct cagtcctcgc gggcctatgt 24180
 cacaccgtcg cgtgggcat tatcgtcgca cagccagaac ccgtccacc gccgcctca 24240
 gaaggcgccc acgcacatt tcaatgcctt tgtcttggg tcgttgagcc gatcaaccg 24300
 ccggttgaat tcgcgtcgc acgcgtcgc accgatggca ttgccaccg ccgcggcgcc 24360
 gtgcacatat gcgttgagc catccccag ttgcgcggac agcgcggcgc tcagactgc 24420
 tgagaccgtc gaggcactgt tgttgagcgc gtcatggcc ggaccttcgc tcggcccggt 24480
 gttgcggccc tgattgaac cgccacgta ggcgttcaac ttgtcatggt cgtctctgt 24540
 ggtggccccc agcgcgtcac acgaggtcgc aatcgcttg gtctcagcg attgttgcg 24600
 ctgcactcc cgtatgtcg acgtcgcgc cgaagccgac accgacgcgc acaccgacga 24660
 gcggtaggcc ggtgcgagct tgggtgcgg catggccgta ccgtcgtga cagtgtaca 24720
 tcgcagcat ccatcagca gcagcgcgt gcagccgag gccaggcgcc ctgcctggg 24780
 gacgtcccc ccgtgctgc gaggcagcgc gcgccatcc atgagcacgg catgtgagg 24840
 tacctggtcg cagcgcgacc gcgctggcgc tgggtgtcgc cgcctccga gaaccgagcg 24900
 gagtgcggct atccgcgcgc gacgccggtg cggcacgata gggggaagac catctaaaca 24960
 gcacgcaagc ggaagccgc cacctacagg agtagtcgt tgaccacga ttctgccgc 25020
 cagcatctg ccaaaaactc aaacagcgca agcgaaccg acgagttcg ggtgatccgc 25080
 gagggtgtg cgtcgtattt gccgcacatt gatccgagg agacctcga gtggctggag 25140
 tctttgaca cgtcgtcga acgctcggc ccgtcgggg ccgctacct gatgttcgg 25200
 ctgctagac gggccgcga gcagcgggtg gccatcccg cattgacgtc taccgatat 25260
 gtcaacacca tcccacgca gctggagccg tggttcccc gcgacgaaga cgtgaacgt 25320
 cgttatcgag cgtggatcag atggaatcgc gccatcatg tgcaccgtc gcaacgacc 25380

ggtgtgggocg tgggtggcca tatctcgacc tacgcgtcgt ccgcggcgct ctatgaggtc 25440
 ggtttcaacc acttcttcgc cggcaagtcg caccggggcg gcggcgatca ggtgttcate 25500
 caggggccacg cttccccggg aatctacgcg cgcgccttcc tccaaggcggt gttgacgcgc 25560
 gagcaactcg acggattccg ccaggaacac agccatgtcg gcggcggtt gccgtctcat 25620
 ccgcaccgcg ggtctatgcc cgaattctcg gaattcccca ccgtgtcgat ggggttgggc 25680
 ccgctcaacg ccatctacca ggcacgggtc aaccactatc tgcattgacc cggatcaaaa 25740
 gacacctccg atcaaacagt gtggtgtttt ttggggcgac gcgagatgga cgaaccgcag 25800
 agccgtgggc tggccacagt cggcgcgctg gaaggcttg acaacttgac cttcgtgatc 25860
 aactgcaatc tgcagcgact cgacggcccg gtgcgcggca acggcaagat catccaggag 25920
 ctggagtcgt tcttcgcgcg tgcgcgctcg aacgtcacta aggtgtgtgt gggcgcggaa 25980
 tgggatgccc tgctgcacgc cgaccgcgac ggtgcgctcg tgaatttaat gaatacaaca 26040
 cccgatggcg attaccagac ctataaggcc aacgacggcg gctacgtgcg tgaccacttc 26100
 ttgcggcgcg acccaacgac caaggcgctg gtggagaaca tgagcgacca ggaatatcgg 26160
 aacctcaaac ggggcggcca cgattaccgc aaggtttacg ccgcctaccg cgcgcgcgtc 26220
 gaccacaagg gacagccgac ggtgactcgt gccaaagcca tcaaggcta cgcgcgtggc 26280
 aagcatttcg aaggacgcaa tgcacccacc cagatgaaaa aactgacctt ggaagacctt 26340
 aaggagtctc gtgacacgca cgggattccg gtcagcgacg ccagcttga agagaatccg 26400
 tacctgcgcg cctactacca cccggcctc aacgcgcccg agattcgtta catgctcgac 26460
 cggcgccggg ccctcggggg ctttgttccc gagcgcagga ccaagtccaa agcgcgtgac 26520
 ctgccgggtc gcgacatcta cgcgccgctg aaaaagggct ctgggaccca ggaggtggcc 26580
 accacocatg cgacgggtcg cacttccaaa gaagtgttgc gcgacaagca gatcggggcg 26640
 cggatagtcg cgatcattcc cgaacgggcc cgcaccttcg ggatggacgc ctggttcccg 26700
 tcgctaaaag tctatacccg caatggccag ctgtataccg cggttgacgc cgacctgatg 26760
 ctggcctaca aggagagcga agtcgggcag atcctgcacg agggcatcaa cgaagccggg 26820
 tcggtgggct cgttcatcgc ggcgcgcacc tcgtatgcga cgcacaacga accgatgate 26880
 cccatttaca tcttctactc gatgttcggc ttccagcgca ccggcgatag cttctggggc 26940
 gcgcgcgacc agatggctcg agggttctgt ctcggggcca ccgcggggcg caccacctgt 27000
 accgggtgag gctcgcaaca cgcgcagcgt cactcgttgc tgctggccgc caccaccccg 27060
 gcggtggttg cctacgaccc ggctctgcgc tacgaaatcg cctacatcgt ggaagcgga 27120
 ctggccagga tgtcggggga gaacccggag aacatcttct tctacatcac cgtctacaac 27180
 gagcgttacg tgcagccgcg ggagccggag aacttcgatc ccgagggcgt gctgcggggg 27240
 atctaccgct atcagcgcgc caccgagcaa cgcaccaaca aggcgcgatc cctggcctcc 27300

ggggtagcga tgcgccggcg gctgcgggca gcacagatgc tggccgccga gtgggatgtc 27360
 gccgcgcagc tgtggtcggt gaccagttag gccgagctaa accgcgacgg ggtggccatc 27420
 gagaccgaga agctccgccca ccccgatcgg ccggcggggc tgcctacgt gacgagagcg 27480
 ctggagaatg ctccggggccc ggtgatcgcg gtgtcggaact ggatgcgcgc ggtcccccag 27540
 cagatccgac cgtgggtgccc gggcacatac ctacagttgg gcacgcgacg gttcggcttt 27600
 tccgacactc ggcccgccgc tcgcgcgtac ttcaacaccg acgcgcgaac ccaggtggtc 27660
 gcggttttgg aggcgttgcc gggcgacggc gagatcgacc catcggtgcc ggtcgcggcc 27720
 gccgcgccagt accggatcga cgacgtggcg gctgcgcccg agcagaccac ggatcccggt 27780
 cccggggcct aacgcggcg agccgaccgc ctttggccga atcttcaga aatctggcgt 27840
 agtttttagg agtgaacgac aatcagttgg ctccagttgc ccgccgagg tcgcgcgtcg 27900
 aactgctgga cactgtgccc gattcgctgc tcggcggtt gaagcagtac tcggccgggc 27960
 tggccaccga gccagtttcg gccatgcaag aacggttgcc gttcttcgcc gacctagaag 28020
 cgtcccagcg ccgcagcgtg cgcctggttg tcgacggcgc cgtggtcaac ttgctcgaat 28080
 ggatgcacga cccgcacagt gacgtcggtc ataccgcgca ggcattcgag ctggtgcccc 28140
 aggatctgac gcgacggatc cgcctgcgcc agaccgtgga catggtgcgg gtcaccatgg 28200
 agttcttcga agaagtcgtg cccctgctcg ccggttcga agagcagttg accgccctca 28260
 cgggtggcat tttgaatac agccgcgacc tggcattcac ccgccccac gccatcgccg 28320
 atcgggccga ggcacgagcg acctgggaca gccggatgga ggccagcgtg gtggacggcg 28380
 tggtaocggg cgacaccggt ccgagctgc tgtcccgggc ggccgcgctg aattgggaca 28440
 ccaccgcgcc gggcaccgta ctggtgggaa ctccggcgcc cggtcacaaat ggtcccaaca 28500
 gcgacggcga cagcgagcgg gccagccagg atgtccgcga caccgcggct cgcacggcc 28560
 gcgctgcgct gaccgacgtg caccgcacct ggtggtggc gatcgtctcc ggcagctgt 28620
 cgccaaccga gaagtctctc aaagacctgc tggcagcatt cgcgcgaccc ccggtggtca 28680
 tcggcccccac ggcccccatt ctgacccggc cgcaccgcag cgtagcgag gcgctctccg 28740
 ggatgaacgc cgtgcggcgc tggcgcgagg ccgcgcggcc cgtgctggct agggaaacttt 28800
 tgcgccaaag gcgctgatg ggcgacgcct cggcgatcgt ggcctcgcat accgacgtga 28860
 tgcggccctc agccgatgcc ggacgcgcgc tcacgcgagc gctagacgca tatctggatt 28920
 gtggcgcgcg gattgaagct tgtgccagaa agttgttcgt tcattccaaac acagtgcggt 28980
 accggctcaa gcggatcacc gacttcacgc ggccgcatcc caccagacca ccgatgcct 29040
 atgtccttcg ggtggcgccc accgtgggtc aactcaacta tccgacgcgc cactgaagca 29100
 tcgacagcaa tgccgtgtca tagattccct cgcgggtcag aggggttcca gcagggggcc 29160
 cggaaagata ccagggggcg cgtcgacggg aaagtgatcc agacaacagg tcgcgggacg 29220

atctcaaaaa catagcttac aggcccggtt tgttggttat atacaaaaac ctaagacgag 29280
 gttcataatc tgttacacog cgcacaaacog tcttcacagt gttctcttag acacgtgatt 29340
 gcgttgctcg caccocggaca gggttcgcaa accgagggaa tgttgctgcc gtggttccag 29400
 ctgcocggcg cagcgggacca gatcgcggcg tggtcgaaag ccgctgatct agatcttgcc 29460
 cggctgggca ccacccgctc gacgcaggag atcacccgaca ccgcggtcgc ccagccattg 29520
 atcgtgcgcg cgaactctgt ggcocaccag gaaotggcgc gccgatgcgt gctgcgcgcg 29580
 aaggacgtca tcgtggcggc ccaactccgtc ggcgaaatcg cggcctaacc aatgcgcggt 29640
 gtgatagcgc ccgacgcgcg cgtgcgcgtc gccgcacccc gcggcgcgca gatggccaag 29700
 gccgcgcgca ccgagccgac cggcatgtct gcggtgctcg gcggcgacga gaccgagggtg 29760
 ctgagtcgcc tcgagcagct cgaacttggtc ccggcaaaac gcaacgccgc cggccagatc 29820
 gtgctgcgcg gccggtgtgc cgcgttgagc aagctcgcgc aagaccgcgc ggccaaggcg 29880
 cgggtgctgc cactgggtgt cgcgcgagcg ttccacaccc agttcatggc gccccgacct 29940
 gacgcgttgc cggcgccgcg gcccaacatc gcaaccgcgc accccacgcg cacgtgtgtg 30000
 tccaacgcgc accgggaagcc ggtgacatcc gcggcgcgcg cgatggacac cctggtctcc 30060
 cagctcaccc aaccgggtgc atgggacctg tgcaccgcga cgcgtgcgcga acacacagtc 30120
 accgcatgac tggagttccc cccgcgggca acgcttagcg gtatcgccaa accggaactt 30180
 cggggggttc cggcacgcgc cgtcaagtc ccgcgagacc tggacagagc ggcaaaccta 30240
 taaccgcgga ctgcgccaga acaaccacat acccgtagt tcgatttgta cacaacatat 30300
 tacgaaggga agcatgctgt gcctgtcact caggaaagaa tcattgccgc tatcgccgag 30360
 atcatcgaag aggtaaccgc tatcgagcgc tcagagatca ccccgagaaa gtctgtctgc 30420
 gacgaactgc acatcgactc gctgtcgatg gtccagatcg ccgtgcagac ccaggacaaag 30480
 tacggcgctc agatcccgca cggaggacct gccggtctgc gtaccgtcgc tgacgttgct 30540
 gcctacatcc agaagctoga ggaagaaaaa ccggaggcgc ctacggcggt gcgcgcgaag 30600
 attgagtcgc agaaccocga tgcggttgc aaogttcagg cgaggcttga ggccgagtc 30660
 aagtgagtc gcctccacc gctaattggc gtttcccag cgttggtgtg accgcgctca 30720
 cagcgacgac gtgcgtctgc ccggacatcg agagcacgtg gaagggtctg ttggccggcg 30780
 agagcggcat ccacgcactc gaagacgagt tcgtcaccaa gtgggatcta gcggtcaaga 30840
 tcggcggtca cctcaaggat ccggtcgaca gccacatggc ccgactcgac atgcgacgca 30900
 tgtctgactg ccagcggatg ggcaagttgc tgggcggaca gctatgggag tcgcgcgcca 30960
 gcccgagggt cgatccagac cgggttcgcg ttgttgctgc caccggtcta ggtggagccg 31020
 agaggattgt cgagagctac gacctgatga atgcgggcgc ccccggaag gtgtcccgc 31080
 tggccgttca gatgatcatg ccaacagggt ccgcggcggt gatcggtctg cagcttgggg 31140

cccgcgccgg ggtgatgacc cgggtgtcgg cctgttcgtc gggctcggaa gcgatcgccc 31200
 acgcgtggcg tcagatcgtg atggcgacg ccgacgtcgc cgtctcggcg ggtgtcgaa 31260
 gacccatcga ggcycctgcc atcgcgcggt tctccatgat cggggccatg tcgacccgca 31320
 acgacgagcc tgagcggggc tcccgccgt tcgacaagga ccgcgacggc tttgtgttcg 31380
 gcgagggcgg tgcgtgatg ctcatcgaga cggaggagca cgccaaagcc cgtggcgcca 31440
 agcgttgtgc ccgattgctg ggtgccggt tcacotcgga cgcctttcat atggtggcgc 31500
 ccgcggccga tgggtgttcgt gccggtaggg cgtgactcgc ctgcgtggag ctggccgggt 31560
 tgtcgcggcg ggacatcgac cagctcaacg cgcacggcac ggcgacgcct atcgcgacg 31620
 ccgcggaggc caacgccatc cgcgtcgcgc gttgtgatca ggcgcgggtg taocgcgcga 31680
 agtctgcgct gggccactcg atcgcgcgcg tcggtgcgct cgagtgggtg ctacgggtgc 31740
 tgacgttcgc cgacggcgct atcccgccga cctggaacta cgagacaccc gatcccgaga 31800
 tcgaccttga cgtcgtcgcc ggcgaaaccg gctatggcga ttaccgctac gcagtcacaa 31860
 actcgttcgg gttcggcgcg cacaatgtg cgcttgccct cgggcgttac tgaagcaga 31920
 catcgccggg ccgcaggccc gaggtggggg tccccccgct tgcggggcg agtcggaccg 31980
 atatggaagg aacgttcgca agaccaatga cggagctggt taocgggaaa gcctttccct 32040
 acgtagtctg caccggcatc gccatgacga ccgcgctcgc gaccgacgcg gagactaagt 32100
 ggaagtgtgt gctggacgcg caaagcggga tcctacgct cgatgaccca ttgctcgagg 32160
 agttcgacct gccagttcgc atcgcgcgac atctgcttga ggaattcgac caccagctga 32220
 cgggatcga actgcgcgg atgggatacc tgcagcggat gtcacaccgt ctgaccggcg 32280
 gcctgtggga aaatgccgcg tcacccgagg tggacaccaa tcgattgatg gtgtccatcg 32340
 gcaccggcct ggttcgcgcc gaggaactgg tcttcagtta cgacgatag cgcgctcgcg 32400
 gaatgaaggc ggtctcgcgc ctgaccgtgc agaagtacat gcccaacggg gcgcgcgcgg 32460
 cggtcggggt ggaacggcac gccaaaggcg gggtagtgac gccggtatcg gcgtgcgcat 32520
 ccggcgccga ggcacatgcc cgtgcgtggc agcagattgt gctgggagag gccgatgccg 32580
 ccattctcgg cggcgtggag accaggatcg aagcgggtgcc catcgccggg ttogctcaga 32640
 tgcgcacgtg gatgtccacc aacaacgacg accccgcggg tgcattccgc ccattcgaca 32700
 gggaccgcga cggctttgtg ttccggcgagg gcggcgccct tctgttgatc gagaccgagg 32760
 agcaccgcaa ggcacgtggc gccaacatcc tggcccggt catggcgccg agcatcacct 32820
 ccgatggctt ccacatgggt gccccggacc ccaacgggga acgcgcggcg catcgcatga 32880
 cgcggcgcat tcagctggcg ggcctcgccc ccggcgacat cgaccaacgc aatgcgcacg 32940
 ccaccggcac ccaggtgcgc gacctggccg aaggcagggc catcaacaac gccttggggc 33000
 gcaaccgacc ggcggtgtac gccccaagt ctgccctcgg ccactcgggt ggcgcggctg 33060

gcgcggtcga atcgatcttg acggtgctcg cgttgccgga tcagggtatc ccgcgcacac 33120
 tgaatctggt aaacctcgat ccgagatcg atttggacgt ggtggcgggt gaaccgcgac 33180
 cgggcaatta ccggtatcg atcaataact cgttcggatt cggcgccac aacgtggcaa 33240
 tcgcttcgg acggtactaa accccagcgt tacgcgacag gagacctcg atgacaatca 33300
 tggcccccga ggcggttgcg gagtcgctcg accccgcgga tcgctgttg cggctgagca 33360
 acttcttcga cgacggcagc gtggaattgc tgcacgagcg tgaccgctcc ggagtgtcgg 33420
 ccgcggcggg caccgtcaac ggtgtgcgca ccacgcgtt ctgcaccgac ggcaccgtga 33480
 tggcgcgcg cgatggcgctg gaggggtgca cgcacatcgt caacgcctac gacactgcc 33540
 tcgaagacca gagtcccatc gtgggcatct ggcattcggg tggtgcccg ctggtcgaag 33600
 gtgtgcgggc gctgcacgcg gtagggcagg tgttcgaagc catgatccgc cgttcggct 33660
 acatcccgca gatctcggtg gtctgtcggt tcgcgcgcgg cggcgccgcc tacggaccgg 33720
 cgttgaccga cgtcgtcgtc atggcgccgg aaagccgggt gttcgtcac gccgcgacg 33780
 tggtgccgag cgtcacccgg gaggaogtgc acatggcttc gctcggctgg ccgagagacc 33840
 accacaagaa gtccgggggtg tgcacatcg tcgccgacga cgaactcgat gcctacgacc 33900
 gtggcgccg gttggtcgga ttgtttctgc agcaggggca ttctgatcgc agcaaggccg 33960
 aggcgcgtga caccgacatc caccgcctgc tgcgggaatc ctgcgcagct gcctacgacg 34020
 tgcgtccgat cgtgacggcg atcctcgatg cggacacacc gttgcagcag ttccaggcca 34080
 attggcgccc gtgatgggtg gtccggctgg gtccgctgtc gggtcgcagc gtgggtgtac 34140
 tggccaacaa ccgcctacgc ctggcgcgct gcctgaactc cgaaagcgca gagaaggcag 34200
 cgcgtttcgt cggcgctgtc gacgcgttgc ggattccgct ggtgggtggtg gtgatgtgc 34260
 cgggtctatc ccccggtgtc gaccaggagt ggggtggcgt ggtgcgccgt ggccccaagt 34320
 tgcgtcacgc gttcgcgagc tgcaccgttc cgcgggtcac gctggtcac ccgaagacct 34380
 acggcggggc atacattcgc atgaactccc ggtcgttgaa ccgcgaccaag gttctccct 34440
 ggcgcgacgc cgaggtcgcg gtgatggcg ctaaggcggc cgtcggcatc ctgcacaaga 34500
 agaagttggc cgcgcctccg gagcacgaac gogaagcgct gcacgaccag ttggccgcgg 34560
 agcatgagcg catcgcggcg ggggtcgaca gtgcgctgga catcggctgtg gtgcagaga 34620
 agatcgaccc ggcgcatact cgcagcaagc tcaccgaggc gctggcgagc gctccggcac 34680
 ggcgcggcg ccacaagaac atccgcctgt agttctgacc gogagagac cgagaatcgc 34740
 acgcgcgagg tcgcgcgcgt gcgattctgc gtctgctcgc cagttatccc cagcgggtgc 34800
 tggtaacacg gaggcgtccc tcgcatgctc ggaagggtgc tacgcagcg ctaacaattc 34860
 tcgagaaggg cggcggggtc gccaccacgg cgcaattgct cacggtcagc acccgccaac 34920
 agctcgacgt ccaagtgaac aacggcgcc tcgttcgcgt ttggtacggg gtctacgcgg 34980

cacaagagcc ggacctgttg ggcgccttgg cggctctcga tgtgttcatg ggggggcaag 35040
ccgtgcgctg tctgggcacc gccgcgcgtg tgtatggatt cgacacggaa aacaccgtcg 35100
ctatccalat gctcgatccc ggagtaagga tgcgggccac ggtcggtctg atggtccacc 35160
aacgcgtcgg tgcggcgctc caacgggtgt caggtcgtct cgcgaccgcg cccgcattga 35220
ctgccgtgga ggtgcgacga cagttgcgcc gccgcggggc gctggccacc ctgcacgcgg 35280
cactacggtc aatgcgctgc gctcgcagtg aaattgaaaa cgcggttgct gagcagcgag 35340
gccgcgcagg catcgtcgcg gcgcgcgaac tcttaccctt cgcgcgcgga cgcgcggaat 35400
cggccattga gagcgaaggc cggctcgtca tgatcgacca cgggctgcgg ttgccccaac 35460
ttcaataccc gatacacggc caccgtgttg aaatgtggcg agtcgacttc gcctggcccc 35520
acatgcgtct cgcgcgcgaa tacgaaagca tcgagtggca cgcgggacgg gcggagatgc 35580
tgcgcgacaa gacacgcctg gccaaagctcc aagagctcgg gtggaogatt gtcccgattg 35640
tcgtcgacga tgtcagacgc gaacccggcc gcctggcggc ccgcacgcgc cgcacacctg 35700
accgcgcgcg tatggccggc tgaccgctgg tgagcagacg cagagtcgca ctgcggcccg 35760
cgcagtcgca ctctgcgtct gctcgcgctc aacggctgag gaactcctta gccacggcga 35820
ctacgcgctc gcgatcccg ggcaaccagc cgaatccgggt ccggcggtcg aggatctcgt 35880
ccacatccag cgcgccctca tgggtccacc cgtattcgaa ctccgccggc gtcacgtcga 35940
tgccgtcgcc gaccggctcg gtgggccgct cacatgtggc ggccggcagc acgttgcccg 36000
cctcgccccc gtaccgcgcc accagcgact cgggcaatcc ggccgccgat ccggggggcg 36060
gccacgggtt cgcgggtcgc ccgatcagcg gcaggttcgc agtgccggac ttccgggctc 36120
gcaggtgtcg cagcgtgatg gcgcgattca gcacatcctc tgccatgtag cggttattcc 36180
tcagcttgcc gccgaccaca ctgatcacgc ccgacggcga ttcaaaaaa gcgtggctac 36240
gcgaaacgtc ggcgggtcgg ccctggacac cagcaccgcc ggtgtcgatt agcggccgca 36300
atccgcgata ggcaccgatg acatccttgg tgccgacgcg cgtccccaat gcgggtgtca 36360
ccgtatccag caggaaactg atctcttcgg aagacggttg tggcacatcg ggaatccggc 36420
cgggtgcgtc ttctcgtggtc agcccgagat agatccggcc cagctgctcg ggcatggcga 36480
acacgaagcg gttcagctca ccggggatcg gaatgttcag cgcggcagtc ggattggcaa 36540
acgaactcgc gtcgaagacc agatgtgtgc cgcggctggg gcgtagcctc agggacgggt 36600
cgatctcacc cgcaccacag ccgcgcgcgt tgatgacggc acgcgcgacg agcgcgaacg 36660
actgcgggtt gcgcggctcg gtcaactcca ccgaagtgc ggtgacattc gacgcgccca 36720
cgtaagttag gatcgggcg ccgtgctggg ccgcggtcgc cgcgacggcc atgaccagcc 36780
ggcgctcgtc gatcaattgc ccgtcgtacg cagcagaccc accgtcgagg ccgtcccgcc 36840
gaacggtggg agcaatctcc accaccctg acgcgggat tcggcgcgat cggggcaacg 36900

```

tcgccgcgg cgtaccgcgt agcaccgcga aagcgtcgcc ggccaggaaa ccggcacgca 36960
ccaacgcgcc cttggtgtga cccatcgacg gcaacaacgg gaccagttgc ggcatggcat 37020
gcacgagatg aggagcggtg cgtgtcatca ggattccgcg ttcgacggcg ctgcgccggg 37080
cgatgccacc gttgccgtcg gccagatagc gcagaccgcc gtgcaccaac ttcgagctcc 37140
agcgctggtt gccgaacgcc agatcatgct ttccaccaa ggccaccgtc agaccgcggg 37200
tggcagcacc taaggcaatg ccaacaccgg taatgccgcc gcctatcacg atgacgtcga 37260
gtgcgccacc gtccggcaggt gcggtcaggt cggcgagcgc acgcgcgcgc ttgagtgcag 37320
ccgagtgagg catcagcaca aatatccgtt cagtgcgtgg gtaagtccgg tggccagcgc 37380
ggcggaatcg aggatcgaat cgacgatgtc cgcggactgg atggtcgact gggcgatcag 37440
caacaccatg gtccgcagtc gacgagcgtc gccggagcgc acaactgccg accgctgcgc 37500
cactgtcagc cggcgcgcca acccctcgat caggacctgc tggctggtgc cgaggcgctc 37560
ggtgatgtac accctggcca gctccgagtg catgaccgac atgatcagat cgtcaccocg 37620
caaccggctg gccaccgcga caatctgctt taccaacgct tcccggtcgt ccccgctcag 37680
gggcacotcc cgcagcacgt cggcgatatg gctggtcagc atggacgcca tgatcgaccg 37740
ggtgtccggc cagcgaaggt atacggtcgg gcggctcacg cccgcgcgcc gggcgatctc 37800
ggcaagtgtc acccggtcca cgcgtaatc gaagacgcag ctgcgcgcgt cccgcaggat 37860
acgaccaccg gtatccgcgc ggtcattact cattgacgc atgtgaata ctgtaacgcg 37920
tgactcaccg cgaggaactc cttccaccga tgaatggga cgcgtgggga gatcccgccg 37980
cggccaagcc actttctgat ggcgtccggt cgttctgtaa gcaggttgtg ggcctagcgg 38040
actcggagca gcccgaaact gacccgcgc aggtgcagct gcgcccgctc gccctgtcgg 38100
gggcagacca                                     38110

```

<210> 25

<211> 2540

<212> DNA

<213> Homo sapiens

```

<400> 25
gaaaagggtg acaagtccca ttttcaagag aagatgactt ttaacagttt tgaaggatct 60
aaaacttgtt tacctgcaga catcaataag gaagaagaat ttgtagaaga gtttaataga 120
ttaaaaactt ttgctaattt tccaagtggt agtcctgttt cagcatcaac actggcacga 180
gcagggtttc ttatacttg tgaaggagat accgtgcggt gctttagtgt tcatgcagct 240
gtagatagat ggcaatatg agactcagca gttggaagac acaggaagt atcccaaat 300

```

tgcagattta tcaacggctt ttatcttgaa aatagtgcc cgcagctac aaattctggt 360
 atccagaatg gtcagtacaa agttgaaaaa tatctgggaa gcagagatca ttttgccotta 420
 gacaggccat ctgagacaca tgcagactat cttttgagaa ctgggcagggt tgtagatata 480
 tcagacacca tatacccgag gaacctgcc algtattgtg aagaagctag attaaagtcc 540
 tttcagaact ggcagacta tgcacaccta acccaagag agttagcaag tgctggactc 600
 tactacacag gtattgtga ccaagtgcag tgcttttgtt gtggtggaaa actgaaaaat 660
 tgggaacctt gtgactgtgc ctggtcagaa cacaggcgac actttcctaa ttgcttcttt 720
 gttttgggcc ggaactctaa tattcgaagt gaatctgatg ctgtgagttc tgataggaa 780
 ttcccaaatt caacaaatct tccaagaaat ccatccatgg cagattatga agcaggatc 840
 tttacttttg ggacatggat atactcagtt aacaaggagc agcttgcaag agctggattt 900
 tatgctttag gtgaaggtga taaagtaaag tgctttcact gtggaggagg gctaactgat 960
 tgggaagccca gtgaagacc ttgggaacaa catgctaagt ggtatccagg gtgcaaatat 1020
 ctgltagaac agaagggaca agaatatata aacaatatc atttaactca ttoacttgag 1080
 gagtgtctgg taagaactac tgagaaaaa ccatcactaa ctagaagaat tgatgatacc 1140
 atcttccaaa atcctatggt acaagaagct atacgaatgg ggttcagttt caaggacatt 1200
 aagaaaaata tggaggaaaa aattcagata tctgggagca actataaact acttgagggt 1260
 ctggttcgag atctagtga tgcacagaaa gacagtatgc aagatgagtc aagtacagact 1320
 tcattacaga aagagattag tactgaagag cagctaaggc gcctgcaaga ggagaagctt 1380
 tgcaaaatct gtatggatag aaatattgct atogtttttg ttcttctgtg acatctagtc 1440
 acttgtaaac aatgtgtga agcagttgac aagtgtccca tgtgctacac agtcattact 1500
 ttcaagcaaa aaatttttat gtcttaactc aactctatag taggcattgt atgtgtttct 1560
 tattaacctg attgaatgtg tgatgtgaac tgactttaag taatcaggat tgaattccat 1620
 tagcatttgc taccaagtag gaaaaaaaa gtacatggca gtgttttagt tggcaatata 1680
 atctttgaat ttcttgattt ttcagggtat tagctgtatt atccattttt tttactgtta 1740
 ttttaattgaa accatagact aagaataaga agcatcatac tataactgaa cacaatgtgt 1800
 attcatagta tactgattta atttctaagt gtaagtgaat taatcatctg gattttttat 1860
 tcttttcaga taggcctaac aaatggagct ttctgtatat aaatgtggag attagagtta 1920
 atctcccaa tcacataatt tgttttgtgt gaaaaggaa taaattgttc catgctggtg 1980
 gaaagataga gattgttttt agaggttggt tgttgtgttt taggattctg tccattttct 2040
 tgtaaaagga taacacagga cgtgtcgcaa atatgtttgt aaagtgattt gccattgttg 2100
 aaagcgtatt taatgataga atactatcga gccaacatgt actgacatgg aaagatgtca 2160
 gagatatgtt aagtgtaaaa tgcaagtggc gggacactat gtatagtctg agccagatca 2220

aagtatgtat gttgttaata tgcatagaac gagagatttg gaaagatata caccaaactg	2280
ttaaatgttg ttctctctcg gggagggggg gattggggga ggggccccag aggggtttta	2340
gaggggcctt ttcactttcg acttttttca tttgtttctg ttccgatttt ttataagtat	2400
gtagaccccg aagggtttta tgggaactaa catcagtaac ctaacccccg tgactatcct	2460
gtgctcttcc tagggagctg tgttgtttcc caccacccac ccttccctct gaacaaatgc	2520
ctgagtgtctg gggcactttg	2540

<210> 26

<211> 103

<212> RNA

<213> Homo sapiens

<400> 26	
agcuccuaua acaaaagucu guugcuugug uuucacauuu uggauuuuccu aaauaaugu	60
ucucuuuuuu gaaaaggugg acaaguccua uuuucaagag aag	103

<210> 27

<211> 28

<212> RNA

<213> Homo sapiens

<400> 27	
ggauuuuccua auauaauguu cucuuuuu	28

<210> 28

<211> 1619

<212> DNA

<213> Homo sapiens

<400> 28	
ccgccagatt tgaatcgcg gacccgttg cagaggtggc ggcggcgcca tgggtgcccc	60
gacgttgccc cctgcttgc agccctttct caaggaccac cgcctctcta cattcaagaa	120
ctgcccttc ttggaggct cgcctgtcac cccggagcgg atggccgagg ctggttcat	180
ccactgcccc actgagaacg agccagactt ggcccagtg ttcttctgct tcaaggagct	240
ggaaggctgg gagccagatg acgaccccat agaggaacat aaaaagcatt cgtccggttg	300

```

cgctttcctt tctgtcaaga agcagtttga agaattaacc cttggtgaat ttttgaaact 360
ggacagagaa agagccaaga acaaaattgc aaaggaaacc aacaataaga agaaagaatt 420
tgaggaaact gcgaagaaag tgcgcctgct catcgagcag ctggctgccca tggattgagg 480
cctctggccg gagctgcctg gtcccagagt ggctgcacca cttccagggt ttattccctg 540
gtgccaccag ccttctctgt ggccoccttag caatgtctta ggaaggaga tcaacatttt 600
caaatagat gtttcaactg tgctcctggt ttgtcttgaa agtggcacc aagggtgcttc 660
tgctgtgca gcgggtgctg ctggtaacag tggtgcttc tctctctctc tctctttttt 720
gggggctcat ttttctgttt ttgattcccg ggcttaccag gtgagaagtg agggaggaa 780
aaggcagtgt ccttttctgt agagctgaca gctttgttcg cgtgggcaga gccttcaca 840
gtgaatgtgt ctggacctca tgttgttgag gctgtcacag tctgtagtgt ggaattggca 900
ggtgcctggt gaattctgag tgcagggttc ttatctgtca cactgtgccc tctcagagg 960
acagtttttt tgttgttgtg tttttttgtt tttttttttt ggtagatgca tgacttgtgt 1020
gtgatgagag aatggagaca gagtccctgg ctctctact gttaacaac atggctttct 1080
tattttgttt gaattgttaa ttcacagaat agcacaaact acaattaaaa ctaagcaca 1140
agccattcta agtcattggg gaaaagggtt gaacttcagg tggatgagga gacagaatag 1200
agtgatagga agcgtctggc agatactct tttgccactg ctgtgtgatt agacaggccc 1260
agtgagccgc ggggcacatg ctggccgctc ctccctcaga aaaaggcagt ggocctaaatc 1320
ctttttaaat gacttgctc gatgtctgtg gggactggct gggctgctgc aggcctgtg 1380
tctgtcagcc caaccttcac atctgtcacg ttctccacac gggggagaga cgcagtcgc 1440
ccaggctccc gctttctttg gaggcagcag ctccgcagg gctgaagtct ggcgtaagat 1500
gatggatttg attgccttc ctccctgtoa tagagctgca ggggtggattg ttacagcttc 1560
gctggaacc tctggaggtc atctcggctg tctctgaga ataaaaagcc tgctatttc 1619

```

<210> 29

<211> 27

<212> RNA

<213> Homo sapiens

<400> 29
ggcgucacac cuucggguga agucgcc

27

<210> 30

<211> 27

<212> RNA

<213> Homo sapiens

<400> 30
ggcgucacac cuucggguga agucgcc

27

<210> 31

<211> 12

<212> PRT

<213> Homo sapiens

<400> 31

Tyr	Gly	Arg	Lys	Lys	Arg	Arg	Gln	Arg	Arg	Arg	Pro
1				5						10	

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US02/11757

A. CLASSIFICATION OF SUBJECT MATTER IPC(7) : C12Q 1/68; C07H 21/02; G01N 27/26 US CL : 435/6; 536/23.1; 204/451 According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) U.S. : 435/6; 536/23.1; 204/451 Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) STN, EAST		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 6329146 B1 (Crouke et al) 11 December 2001 (11.12.2001), column 40, example 11.	1
Y	US 5,807,682 A (Grossman et al) 15 September 1998 (15.09.1998), column 19, lines 2-18.	1
Y	US 6,355,428 (Schroth et al) 12 March 2002 (12.03.2002), column 8, lines 64-67.	1
Y	US 6,320,040 B1 (Cook et al) 20 November 2001 (20.11.2001), column 11, lines 14-22	1
Y	US 6,391,542 B1 (Anderson et al) 12 May 2002 (12.05.2002), column 36, example 18.	1
<input type="checkbox"/> Further documents are listed in the continuation of Box C. <input type="checkbox"/> See patent family annex.		
* Special categories of cited documents:		
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"
"B"	earlier application or patent published on or after the international filing date	"X"
"L"	document which may throw doubt on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"Y"
"O"	document referring to an oral disclosure, use, exhibition or other means	"A"
"P"	document published prior to the international filing date but later than the priority date claimed	"A"
Date of the actual completion of the international search 22 June 2002 (22.06.2002) Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703)305-3230		Date of mailing of the international search report 18 JUL 2002 Authorized officer Jyothana Venkat Jyothana Venkat Ph.D Telephone No. (703) 305-3235 EXAMINER TECHNOLOGY CENTER 1600